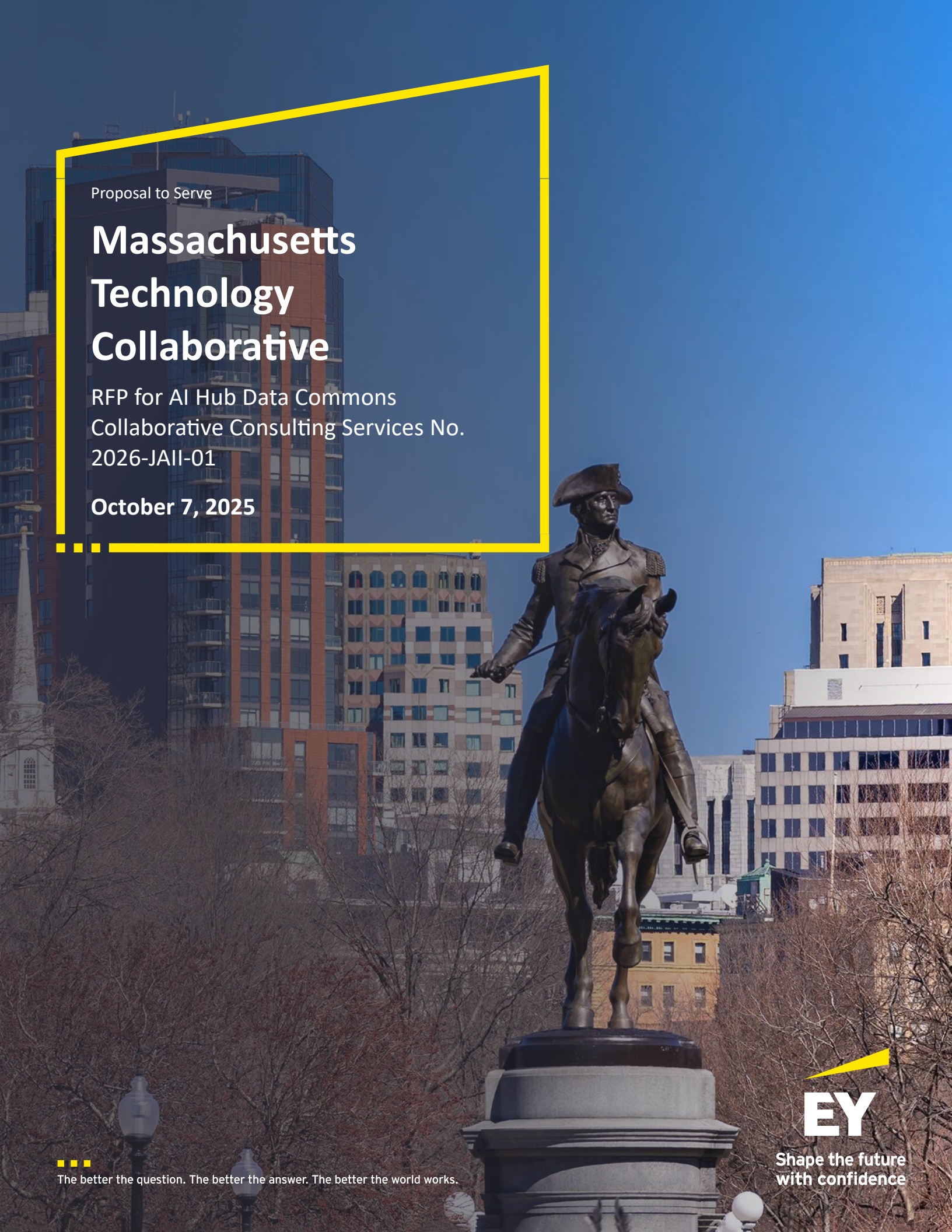


Proposal to Serve

Massachusetts Technology Collaborative

RFP for AI Hub Data Commons Collaborative Consulting Services No. 2026-JAII-01

October 7, 2025

The better the question. The better the answer. The better the world works.

Shape the future with confidence

Shape the future
with confidence

Ernst & Young US LLP
200 Clarendon Street
Boston, MA 02116

Tel: +1 617 266 2000
Fax: +1 617 266 5843
ey.com

Attachment A

Application Cover Sheet

Name of Respondent: Ernst & Young US LLP (EY)			
Mailing Address: 200 Clarendon Street	Boston	Massachusetts	02116
Telephone: 617-921-3101	**Fax**: 617-266-5843	**Web Address:** https://www.ey.com	
Primary Contact for Clarification: Josh Liberman		**Primary Contact E-mail Address**: josh.liberman@ey.com	
Authorized Signatory: Josh Liberman		**Authorized Signatory E-mail Address**: josh.liberman@ey.com	
Legal Status / Jurisdiction: A Delaware registered LLP		**Respondents W-9:** W9 document attached to submission **Respondents UEI No.:** ECMMFNMSLXM7	

Shape the future
with confidence

Ernst & Young US LLP
200 Clarendon Street
Boston, MA 02116

Tel: +1 617 266 2000
Fax: +1 617 266 5843
ey.com

October 7, 2025

Massachusetts Technology Collaborative
75 North Drive
Westborough, MA 01581

Subject: Response to RFP No. 2026-JAII-01 – AI Hub Data Commons Collaborative Consulting Services

Dear Selection Committee,

On behalf of Ernst & Young US LLP (EY), we are pleased to submit our proposal to support the Massachusetts Technology Collaborative AI Hub (MassTech AI Hub) in advancing the Data Commons Collaborative initiative. EY is committed to helping the Commonwealth of Massachusetts realize its vision for a secure, scalable, and innovative data platform that empowers research, collaboration, and responsible AI development.

EY fully recognizes the complexity and ambition of the solution that the MassTech AI Hub is seeking for the Data Commons Collaborative (DCC).

The vision for the DCC is to accelerate AI innovation in Massachusetts by centralizing access to curated and synthetic datasets, enabling research, development, and equitable commercial growth in AI across the Commonwealth. The DCC will be an extensible platform for researchers, government, startups and industry, ensuring ethical, fair and secure AI development, furthering the Commonwealth's goal to become a global leader in applied AI innovation.

The goals of this project are noble, purposeful, and forward-thinking. And although the end goals and high-level requirements are articulated in the RFP and supporting documents, an endeavor of this magnitude warrants detailed, thoughtful planning that includes: clearly defined and measurable goals; specific and detailed functional requirements; technology component selection that balances compatibility, capability, and predictable costs; a secure, extensible and adaptable infrastructure; experience design that meets accessibility standards; as well as an actionable, achievable and phased implementation plan complete with a backlog of tasks across all workstreams so that at the end of this engagement the MassTech AI Hub is equipped with an immediately actionable implementation plan.

To arrive at the necessary level of specificity and actionable detail, EY is proposing a short-duration, focused Discovery, Design, and Planning engagement to refine an architecture and functional capability set that not only achieves the stated goals that the MassTech AI Hub has shared, but establishes a modular foundation that will accommodate functional and technical growth for years to come with predictable costs and an achievable timeline.

Throughout our response you will see the depth to which we have unpacked all the necessary workstreams and functional and technical areas that require attention to arrive at the detail necessary for an actionable, prioritized implementation plan, one that is informed by real-world constraints, leading practices, and a shared understanding of success, that the MassTech AI Hub will rely on for the next 5 years. This approach will lay the groundwork for a successful implementation, with a comprehensive understanding of cost and effort, to realize the full potential of the Data Commons Collaborative.

Within our response you will see two proposed project durations, 16-weeks and 12-weeks, to provide flexibility and align with the pace that MassTech AI Hub can realistically accommodate. Both approaches are identical in scope and outcomes, however, we understand MassTech AI Hub wants to move quickly, and we prepared to match the pace that the MassTech AI Hub can accommodate. Please note, the 12-week option is a lower investment for the MassTech AI Hub with the same outcome.

With over 2,300 professionals living and working in Massachusetts and a proven track record of successful engagements with Commonwealth state agencies, EY is uniquely positioned to deliver rapid, high-quality results. We have assembled a highly skilled and experienced team, drawing on deep expertise from EY's Technology Consulting practice. This collaborative team brings together leading practitioners in hybrid cloud, AI, data engineering, and security, supported by proven project leadership methodologies and a track record of successful public sector transformations.

Our commitment is to work side-by-side with the MassTech AI Hub, leveraging our collective strengths to define a solution that is both innovative and sustainable. We look forward to the opportunity to collaborate with your team and contribute to the Commonwealth's leadership in AI and data innovation.

Shape the future
with confidence

Ernst & Young US LLP
200 Clarendon Street
Boston, MA 02116

Tel: +1 617 266 2000
Fax: +1 617 266 5843
ey.com

Thank you for considering our proposal. Please do not hesitate to contact us if you require any additional information or clarification.

Sincerely,

Josh Liberman
Managing Director
Government & Public Sector
Lead Client Partner for Commonwealth of Massachusetts

Matt Allen
Managing Director
Technology Consulting
Government & Public Sector

Table of contents

Description of Firm Capabilities & Experience 1

Technical Expertise 1

Platform Delivery Experience 3

Familiarity with Metadata Standards 4

Familiarity with Synthetic Data 6

Familiarity with AI Fairness Tools 8

EY's Alliance Ecosystem 10

Approach 13

Overview 14

Detailed Project Approach and Plan 14

High-Level Technical Architecture Considerations 36

Resumes & Bios 38

Qualifications and availability of proposed team 38

References & Quals 42

Strength and Relevance of References and Prior Work 42

References 43

Cost Response (Attachment C) 45

Budget realism, alignment with scope, and cost-effectiveness 45

W-9 47

Signature Acceptance Form (Attachment B) 48

Appendix A – Proposal Assumptions 49

General Project Assumptions: 49

Scope Assumptions: 49

Appendix B – Compressed Timeline 51

12-Week Timeline 51

12-Week Activity Schedule 51

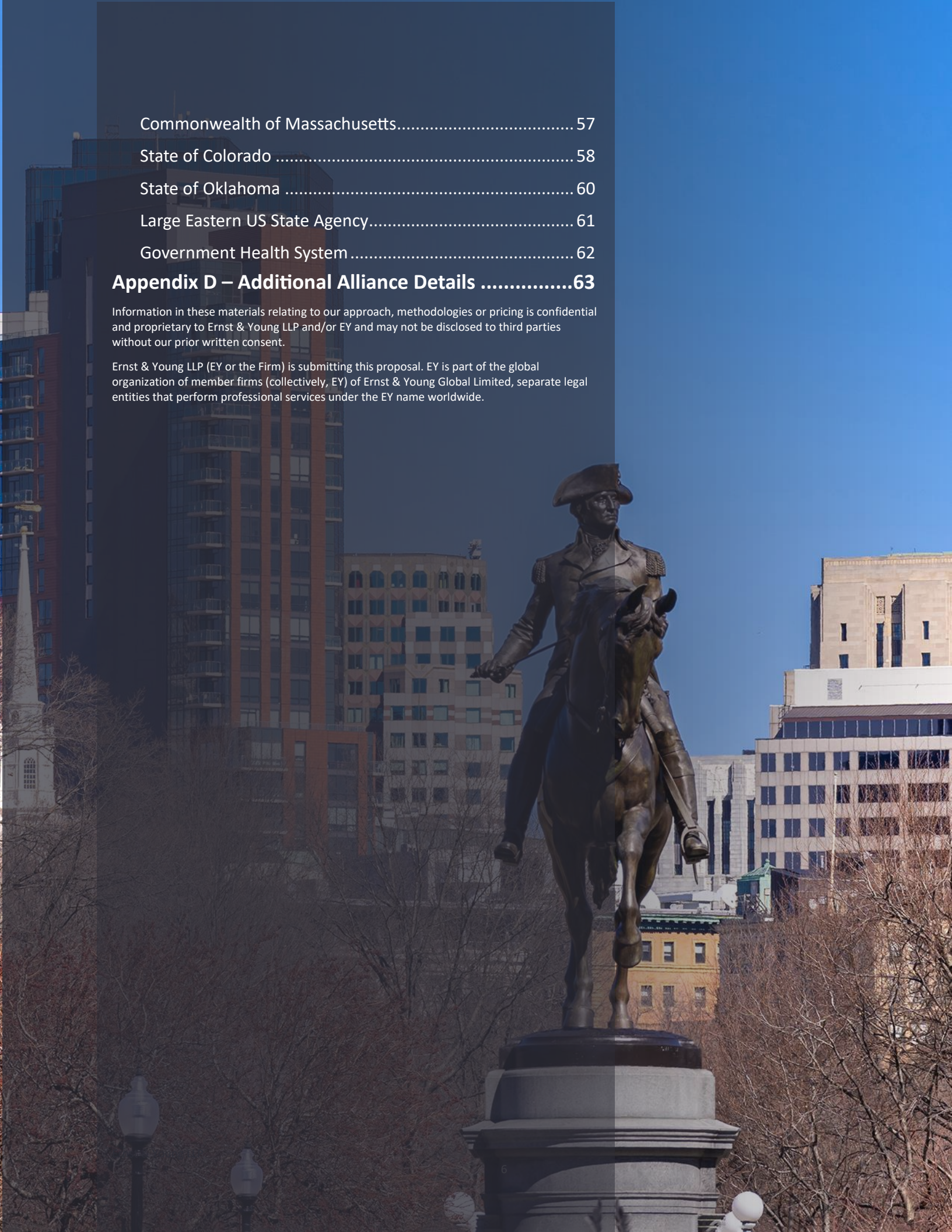
Appendix C - Related Experience 57

Commonwealth of Massachusetts 57

Commonwealth of Massachusetts..... 57
State of Colorado 58
State of Oklahoma 60
Large Eastern US State Agency..... 61
Government Health System 62
Appendix D – Additional Alliance Details 63

Information in these materials relating to our approach, methodologies or pricing is confidential and proprietary to Ernst & Young LLP and/or EY and may not be disclosed to third parties without our prior written consent.

Ernst & Young LLP (EY or the Firm) is submitting this proposal. EY is part of the global organization of member firms (collectively, EY) of Ernst & Young Global Limited, separate legal entities that perform professional services under the EY name worldwide.



Description of Firm Capabilities & Experience

EY has deep and successful experience innovating in AI strategy and implementation for organizations large and small around the world. EY was named as a global leader in IDC's MarketScape Worldwide Artificial Intelligence Services rankings this year. EY has done similar work of data estate and AI solution design and implementation with agencies in Massachusetts, Virginia and Louisiana among many other states.

EY has over 400,000 professionals worldwide delivering consulting services across several service lines including Technology Consulting, and with over 2,300 professionals living and working in Massachusetts, we are positioned to rapidly respond to client needs across a broad spectrum of capabilities.

EY is committed to serving the Commonwealth of Massachusetts with over 90+ EY partners serving on boards and volunteering time with Boys and Girls Club of Boston, Earthwatch Institute, the Massachusetts Mentoring Partnership, the Massachusetts Opportunity Alliance, the Boston Public Library, the JFK Library and more. EY has also provided over $1 million dollars in community engagement funding to 76 Massachusetts non-profit organizations. EY works, lives, and builds community in Massachusetts.

Our work within the Commonwealth spans more than 20 state agencies, academic institutions, and the State Legislature. We understand the challenges faced by the state government in Massachusetts and are proud to be working alongside many state agencies to transform and innovate along with them.

MassTech can draw on our extensive experience serving clients in the public sector as well as in the commercial space, including 36 of the 50 States in the Union, 19 Federal agencies, and 96% of Fortune 50 companies. Much of this work has been on the exact workstreams in scope of this solicitation e.g., architecture design, metadata management, security and governance, AI innovation and fairness.

Beyond just numbers, EY's Technology Consulting service line brings deep capabilities in AI and data-related consulting services from responsible AI, data infrastructure to data engineering, to modern data estate design and implementation. Our alliances with leading solution providers such as Snowflake and IBM allow us to bring the very best that the market has to offer and allows our team members to stay up to date on all current trends and technology capabilities.

Technical Expertise

EY brings a deep bench of architects, data engineers, AI/ML practitioners, privacy and security specialists, and platform engineers with hands-on experience designing, building, and operating secure, multi-tenant data and AI platforms for public sector and research environments. Our technical capabilities align directly to the Data Commons Collaborative (DCC) vision and the requirements of the MassTech AI Hub.

Core platform architecture and engineering

- Hybrid cloud reference architecture: Design and implementation of an extensible, cloud-agnostic platform leveraging Red Hat OpenShift for container orchestration, Kubernetes-native services, and cloud services across providers to support elastic compute (CPU/GPU), secure storage, and high-throughput data processing.
- Data lakehouse pattern: Lake and warehouse layers built on columnar formats (Parquet) and open table formats (Iceberg/Delta) with governance and analytics through Snowflake for secure data sharing, native apps, and workload isolation. Support for streaming ingestion via Kafka and event-driven pipelines.
- Research workspaces: Provisioning of secure, isolated workspaces (e.g., JupyterLab, RStudio, VS Code) with hardened images, role-based access, and policy controls to enable reproducible research while protecting sensitive data.
- API and integration fabric: Standards-based REST/GraphQL APIs, data virtualization, and service mesh patterns to enable interoperability with agency systems, universities, and industry organizations.

Data onboarding, curation, and metadata/lineage

- Ingestion pipelines: ELT/ETL pipelines with automated schema discovery, data profiling, quality rules, and observability (data freshness, completeness, anomaly detection) using modern frameworks and orchestration tools.
- Metadata and catalog: Enterprise cataloging with support for DCAT and schema.org Dataset metadata, business glossaries, and dataset/data card generation to document provenance, intended use, and limitations.
- Lineage and reproducibility: End-to-end lineage based on OpenLineage standards, versioning of datasets and transformations, and reproducible pipelines to support auditability and trust.
- Stewardship workflows: Data access request and approval workflows, license/terms-of-use tagging, and automated enforcement of use conditions.

Privacy-preserving data and synthetic data capabilities

- De-identification and anonymization: Expert implementation of k-anonymity, l-diversity, t-closeness, and tokenization, combined with risk scoring and re-identification testing to protect individuals while retaining utility.
- Differential privacy: Calibrated noise mechanisms for statistical releases and model training that balance privacy budgets with analytical value.
- Synthetic data factory: Generation of high-fidelity synthetic datasets across tabular, time-series, and unstructured modalities using generative techniques. Utility validation (statistical similarity, downstream model performance) and privacy safety tests (membership inference, attribute inference) to promote responsible use.
- Secure data collaboration: Federated analytics and model training patterns for scenarios where data cannot be moved, enabling cross-institution collaboration with privacy controls.

AI/ML engineering, MLOps, and Responsible AI

- MLOps at scale: CI/CD for models and data pipelines, feature stores, experiment tracking, model registry, and automated deployment with canary strategies, rollback, and drift detection.
- Governance and risk management: Alignment to NIST AI RMF 1.0 and emerging ISO/IEC AI management standards; bias and fairness assessments, interpretability, robustness testing, and human-in-the-loop review.
- Model and data documentation: Model cards, data cards, and transparency artifacts integrated into the catalog to support reproducibility and ethical use.
- Monitoring and quality: Real-time telemetry for model performance, data drift, and fairness metrics, with alerts and playbooks for remediation.

Security, compliance, and policy enforcement

- Zero Trust architecture: Strong identity and access management, multi-factor authentication, and fine-grained, attribute-based access control (ABAC) with policy-as-code to enforce data-use restrictions.
- Encryption and key management: Encryption in transit and at rest, HSM-backed key management, secrets management, and secure enclave options for sensitive workloads.
- Auditability: Comprehensive logging, immutable audit trails, and evidence collection to support compliance with state and federal requirements (e.g., HIPAA, FERPA, CJIS as applicable).
- Data governance: Classification, tagging, retention, and legal hold processes; automated enforcement of licensing and data-use agreements across datasets and derived assets.

Performance, scalability, and cost management

- Elastic compute and GPU orchestration: Capacity planning for AI workloads, automated scaling, and queueing for shared GPU clusters supporting advanced research and training.
- High-performance data access: Storage and caching strategies for large datasets, parallel I/O, and query acceleration.
- FinOps: Cost visibility, budgeting, and guardrails at the dataset, project, and workspace level to promote sustainable operations.

Alliances and technology ecosystem

- IBM: Hybrid cloud and AI solutions including Red Hat OpenShift for portable, secure compute and watsonx capabilities for model governance to accelerate compliant AI adoption.
- Snowflake: Cloud-native data platform for secure data sharing, governance, workload isolation, and Snowpark for scalable data engineering and ML close to data.
- Open-source and standards: Pragmatic use of proven open-source technologies and adherence to open standards to avoid lock-in and enhance interoperability across the Commonwealth.

Practitioner depth and delivery approach

- Specialized roles: Solution architects, data engineers, ML engineers, privacy engineers, security architects, platform/SRE engineers, and product managers with public-sector experience and certifications across cloud, security, and data platforms.
- Proven methods: Discovery, Design, and Planning focused on stakeholder alignment, risk-managed architecture choices, and executable roadmaps. We bring accelerators for cataloging, data quality, privacy testing, MLOps, and Responsible AI to shorten time-to-value.

This technical foundation enables MassTech AI Hub's DCC to centralize curated and synthetic datasets, provide secure and equitable access for researchers and industry, and operationalize responsible AI at scale while remaining modular for future growth and adaptable to evolving policy and technology landscapes.

Platform Delivery Experience

EY delivers complex data and AI platforms using a "platform as a product" operating model that emphasizes reliability, security, and rapid value delivery for multi-stakeholder environments like the DCC. Our approach blends proven public-sector delivery methods with modern DevSecOps, MLOps, and Site Reliability Engineering (SRE) practices to stand up, scale, and sustain the platform.

Platform-as-a-Product operating model

- Dedicated product leadership: Product Owner, Platform Architect, and Service Manager roles to align releases with MassTech AI Hub priorities and stakeholder needs.
- Backlog and intake: Structured demand management for dataset onboarding, workspace requests, feature development, and policy updates; prioritization based on public value, risk, and readiness.
- Service catalog: Clear, requestable services (e.g., dataset onboarding, workspace provisioning, GPU quota, synthetic data generation, model review) with defined SLAs and eligibility criteria.

Delivery methods and governance

- Iterative delivery: Discovery, Design, and Planning followed by MVP releases and progressive capability expansion; each increment gated by security, privacy, and performance readiness checks.
- Multi-stakeholder governance: Steering committee and user councils (researchers, agencies, academia, startups) to promote transparent decision-making, equitable access, and responsible AI guardrails.
- Risk-managed execution: Early identification and mitigation of data licensing, privacy, compute, and interoperability risks; dependency tracking across agencies and organizations.

Environment build and infrastructure operations

- Hybrid cloud foundation: Portable container orchestration (e.g., OpenShift/Kubernetes), secure networking, workload isolation, GPU scheduling, and storage tiers configured for large datasets.
- Infrastructure as Code: Automated provisioning, configuration baselines, and environment parity across dev/test/prod to improve repeatability and reduce drift.
- Performance engineering: Capacity planning for CPU/GPU, caching and query acceleration, and high-throughput data I/O to support research and training workloads.

DevSecOps, DataOps, and MLOps

- CI/CD pipelines: Automated build, test, security scanning, SBOM generation, container signing, and policy-as-code enforcement for platform services and data pipelines.
- DataOps: Standardized ingestion, schema evolution controls, quality checks, lineage capture, and observability to promote trusted datasets.
- MLOps: Experiment tracking, feature stores, model registry, approval workflows, deployment strategies (canary/blue-green), and monitoring for drift, bias, and performance.

Security, privacy, and compliance operations

- Zero Trust: Strong identity, multi-factor authentication, and fine-grained, attribute-based access with policy enforcement aligned to Commonwealth standards.
- Privacy-by-design: Differential privacy options, de-identification controls, re-identification risk testing, and enforceable license/terms-of-use tagging.
- Continuous compliance: Immutable audit trails, evidence collection, vulnerability management, patching cadence, and support for applicable regulatory frameworks (e.g., HIPAA, FERPA, CJIS as relevant to datasets).

SRE, observability, and resilience

- SLOs and SLAs: Reliability targets for core services (e.g., catalog availability, workspace provisioning, pipeline success rates) with error budgets and incident playbooks.

- End-to-end observability: Telemetry across infrastructure, data pipelines, and models for proactive detection and rapid remediation.
- Business continuity: Backup/restore, cross-region failover strategies, and disaster recovery procedures tailored to research workloads and data sensitivity.

Data onboarding and collaboration operations

- Intake and curation: Standardized workflows for dataset intake, metadata enrichment (DCAT/schema.org), provenance documentation, and stewardship assignment.
- Synthetic data factory: Operational processes to generate, validate, and publish synthetic datasets with utility and privacy safety tests; transparent documentation and labeling.
- Federated collaboration: Secure patterns for cross-institution analytics when data cannot move, enabling participation by agencies, universities, and industry.

Service management and adoption

- ITIL-aligned processes: Incident, problem, change, and configuration management with a knowledge base and runbooks.
- Enablement and training: Role-based training (researchers, stewards, platform operators), onboarding toolkits, and office hours to drive adoption and responsible use.
- Change management: Release notes, stakeholder briefings, and impact assessments to minimize disruption and build trust across the ecosystem.

FinOps and cost stewardship

- Cost visibility: Tagging and reporting at dataset, project, and workspace levels; budget guardrails and alerts.
- Cost Control: Workload right-sizing, GPU scheduling policies, storage tiering, and architectural patterns that balance performance and cost.

Alliances and accelerators

- IBM and Snowflake: Pre-validated reference architectures and governance patterns to accelerate secure data sharing, workload isolation, and responsible AI operations.
- EY accelerators: Cataloging, privacy testing, data quality rules, MLOps templates, and governance playbooks to shorten time-to-value and reduce delivery risk.

Transition and sustainment

- Handover and capability uplift: Documentation, standardized runbooks, and coaching to transition operations to MassTech AI Hub teams or designated operators.
- Continuous improvement: Quarterly roadmap reviews, benchmarking, and feedback loops to evolve the platform with emerging standards and stakeholder needs.

This disciplined, proven delivery approach enables MassTech AI Hub to stand up the DCC quickly, onboard diverse data providers, support secure research workspaces at scale, and maintain a high-trust environment for responsible AI innovation across the Commonwealth.

Familiarity with Metadata Standards

EY has deep, practical experience designing and operating metadata frameworks that make data findable, accessible, interoperable, and reusable (FAIR) across public-sector and research ecosystems. For the DCC, we will establish a standards-based, extensible metadata backbone that supports discovery, provenance, licensing, access control, and reproducibility across curated and synthetic datasets.

Foundational dataset discovery and catalog standards

- W3C DCAT as the canonical backbone, with support for common profiles (e.g., Project Open Data/DCAT-US used by data.gov, DCAT-AP-style profiles) to enable publication and federation with external portals.
- schema.org/Dataset and JSON-LD markup to improve discoverability on the open web and to enable lightweight interop with academic and industry organizations.
- Dublin Core for core descriptive elements where appropriate and for crosswalks with legacy catalogs.

- Persistent identifiers and registries: DataCite DOIs for datasets and versions, ORCID/ROR for people and organizations, ensuring citation, attribution, and durable links.

Provenance, lineage, and reproducibility

- W3C PROV-O for representing provenance relationships among datasets, processes, and agents; used to capture source, transformations, and responsible parties.
- OpenLineage for operational lineage across pipelines and tools (e.g., Airflow, Spark, dbt), enabling end-to-end traceability and impact analysis.
- Research object packaging: RO-Crate for portable, citable bundles that include data, code, environment specs, and documentation for reproducible research.

Domain-specific metadata and ontologies

- Geospatial: ISO 19115/19139 for dataset description, OGC standards (WMS/WFS/WCS) and STAC for spatiotemporal asset catalogs (particularly for remote sensing/EO data).
- Health and biomedical: HL7 FHIR resource profiles, OMOP Common Data Model, CDISC (SDTM/ADaM), and NIH Data Tag Suite (DATS) to document clinical/observational datasets and enable responsible secondary use.
- Justice/public safety and human services: NIEM for interagency interoperability of structured entities and relationships.
- Education: Common Education Data Standards (CEDS), Ed-Fi, and IMS Global profiles to describe education data consistently.
- IoT/sensors and time series: OGC SensorThings API and SensorML for sensor, observation, and temporal metadata.
- Controlled vocabularies and semantics: SKOS for vocabularies; integration with authoritative thesauri/ontologies (e.g., MeSH, SNOMED CT, EPSG codes) to standardize subjects, units, and georeferencing.

Licensing, usage rights, and policy metadata

- SPDX license identifiers, Creative Commons, and Open Data Commons to unambiguously declare dataset licenses and attribution requirements.
- ODRL (Open Digital Rights Language) and machine-readable terms-of-use to encode allowed/forbidden uses, redistribution constraints, and citation expectations.
- Sensitivity and classification tags aligned to Commonwealth policy (e.g., public/restricted/confidential), NIST guidance, and HIPAA/FERPA contexts, enabling attribute-based access control (ABAC) and policy-as-code enforcement.

Data quality, tabular documentation, and FAIR enablement

- Frictionless Data specifications (Data Package, Table Schema) and W3C CSV on the Web (CSVW) for schema, constraints, and rich column-level annotations on tabular datasets.
- Dataset quality and readiness indicators (freshness, completeness, validity) embedded as metadata to drive trust and automated checks in pipelines.
- Multilingual metadata and accessibility tags to support equitable discovery and use.

AI/ML and synthetic data documentation

- Datasheets for Datasets and Data Cards to standardize documentation of purpose, collection methods, populations, known biases, and intended uses/limitations.
- Model Cards and IBM AI FactSheets-style documentation for model governance; linkage to training datasets via PROV-O and to feature stores via standardized identifiers.
- Synthetic data metadata: generator type and version, training sources, privacy parameters (e.g., epsilon for differential privacy), utility metrics, and results of privacy safety tests (e.g., membership/attribute inference), clearly labeled and linked to original datasets and licenses.

Implementation approach and tooling

- Canonical metadata model: Establish a DCAT-based core with DCC-specific extensions and mappings to domain profiles; publish as an open schema and versioned profile.
- Crosswalks and adapters: Automated mappings to/from CKAN/DKAN, Collibra/Alation, IBM Knowledge Catalog, Snowflake object tags, Apache Atlas/Egeria, OpenMetadata/DataHub, ensuring the platform does not lock into a single vendor model.

- Encodings and APIs: JSON-LD/RDF for semantic interoperability; DCAT feeds for external publication; STAC APIs for geospatial collections; GraphQL/REST for programmatic access; harvesting and synchronization jobs for external catalogs.
- Automated enrichment: PII detection/classification, schema inference, data profiling, and lineage capture integrated into ingestion pipelines to keep metadata current and reliable.

Governance and stewardship

- Metadata stewardship model with clear ownership for datasets, collections, and vocabularies; change control for schema/profile evolution.
- Metadata quality controls and certification levels (e.g., bronze/silver/gold) to signal readiness for research or production.
- Persistent identifier workflows for dataset versions and citations; citation guidance embedded in dataset pages.

What this means for the DCC

- We will adopt DCAT as the backbone with a DCC profile, extend for synthetic data and responsible AI artifacts, and support domain profiles (e.g., STAC, FHIR/OMOP) where needed.
- We will enable machine-readable licenses and policy tags that drive automated access control and responsible-use enforcement across research workspaces.
- We will implement provenance and lineage end-to-end so researchers can reproduce results and regulators can audit usage.
- We will federate with external catalogs and state portals using standard feeds, ensuring the DCC is interoperable by design.

This standards-first approach aligns with goals for DCC datasets which are consistently described, easily discoverable, governed by machine-actionable policies, and portable across institutions, tools, and clouds, supporting secure, equitable, and responsible AI innovation across the Commonwealth.

Familiarity with Synthetic Data

EY has extensive, hands-on experience designing, generating, validating, governing, and operationalizing synthetic data for public-sector, healthcare, financial services, and research use cases. We help clients use synthetic data to unlock collaboration and innovation when real data is sensitive, sparse, or encumbered by licensing, while preserving privacy, fairness, and utility.

Where and why we apply synthetic data

- Privacy-preserving data sharing: Enable cross-agency and academic collaboration when direct access to identifiable datasets is not permissible.
- Safe sandboxes and hackathons: Provide realistic, low-risk datasets for innovation challenges, training, and vendor evaluations.
- Rare event amplification: Increase representation of low-frequency but critical cases (e.g., fraud, outages, adverse events) for robust model development.
- Test data and DevSecOps: Populate lower environments with realistic data that respects data minimization and privacy policies.
- What we don't do: We treat synthetic data as one tool among many. Where high-precision analyses depend on rare signals or causal relationships, or where policy prohibits derivative data release, we recommend alternatives (e.g., secured enclaves, federated analytics, controlled access to de-identified data).

Generation methods we employ

- Statistical synthesis: Copula-based methods and probabilistic graphical models for tabular data with strong constraint handling and explainability.
- ML-based generators: GANs, VAEs, autoregressive/tabular transformers, diffusion models, and time-series generators (e.g., TimeGAN) for complex distributions and sequences.
- Differentially private generation: Train generators with privacy controls using DP-SGD and related mechanisms to manage privacy budgets.
- Hybrid and rule-driven synthesis: Combine learned distributions with domain rules, business logic, and constraints to promote validity (e.g., date ranges, code sets).
- Simulation-based approaches: Agent-based or process simulations where mechanisms are well-understood (e.g., transportation flows), producing fully non-identifying data by design.

Utility and fidelity validation

- Statistical similarity: Compare marginals, joint distributions, pairwise correlations, mutual information, and distance measures (e.g., Wasserstein/KL) against source data.
- Downstream task performance: Train models on synthetic and evaluate on held-out real data; assess accuracy, calibration, and stability.
- Constraint and semantics checks: Validate domain constraints, code-list adherence, and cross-field logic; profile data quality (completeness, validity, uniqueness).
- Representativeness and coverage: Check class balance, tail behavior, and rare pattern fidelity to avoid erasing important minorities or fabricating implausible patterns.

Privacy and leakage risk assessment

- Re-identification resistance: Nearest-neighbor distance analyses, uniqueness checks, and record linkage tests against the source to detect memorization.
- Adversarial testing: Membership and attribute inference attacks to probe leakage; exposure analysis for text/unstructured modalities.
- Differential privacy accounting: Track and report privacy budgets (epsilon/delta) when DP is applied; enforce budget policies at dataset and program levels.
- Policy classification: Risk-tier synthetic outputs (e.g., public synthetic, restricted synthetic) based on measurable privacy risk and intended use.

Fairness and responsible AI considerations

- Bias preservation and mitigation: Evaluate distributions and downstream model fairness metrics (e.g., demographic parity difference, equalized odds) on synthetic vs. real data.
- Transparent documentation: Publish Data Cards/Datasheets for synthetic datasets capturing purpose, generation method, training sources, limitations, and known risks.
- Clear labeling and provenance: Unambiguous “synthetic” flags, provenance links to source datasets (not accessible by default), and license/terms reflecting permitted uses.

Governance, standards, and metadata

- Standards-aligned metadata: Extend DCAT with synthetic-specific fields (generator type/version, training data references, DP parameters, privacy/utility test results).
- Policy-as-code: Encode allowed uses and restrictions (e.g., research-only, no linkage attempts) and enforce via ABAC and automated checks in workspaces.
- Compliance context: Methods aligned to HIPAA de-identification principles, FERPA considerations, and NIST privacy guidance; expert determination workflows where needed.

Synthetic Data Factory for the DCC

- Intake and assessment: Triage candidate datasets, determine whether synthetic is appropriate, define intended uses, and select generation approach.
- Generation and hardening: Implement pipelines in secure environments with versioned code, seeded randomness control, and reproducible containers.
- Validation gates: Automated and human-in-the-loop gates for utility, privacy, and fairness; failure paths trigger retraining or method changes.
- Documentation and publishing: Create dataset cards, attach machine-readable licenses and policy tags, and publish to the DCC catalog with search facets and lineage.
- Post-release monitoring: Track feedback, re-run leakage and drift tests on a cadence, and deprecate/update synthetic datasets as sources evolve.

Tooling and platform integration

- Open frameworks: Experience with SDV (CTGAN/TVAE), copulas, TimeGAN, and constraint solvers; integration with notebook-based workflows and pipeline orchestrators.

- Cloud/data platforms: Run synthetic generation close to data in secure environments; store outputs with clear labeling in Snowflake or object storage with governance tags.
- Federated and on-prem options: For highly sensitive sources, generate models on-premises and export only synthetic outputs and validation artifacts to the DCC.

Illustrative outcomes we have delivered

- Healthcare and human services: Synthetic EHR/claims-like datasets to enable collaboration without exposing PHI, with documented DP budgets and expert determination.
- Tax, labor, and benefits administration: Realistic, policy-constrained tabular and time-series data for fraud analytics development and vendor bake-offs.
- Transportation and geospatial: Synthetic trip and sensor datasets preserving flow patterns while obfuscating individual trajectories, validated with STAC-aligned metadata.

What this means for MassTech AI Hub's DCC

- The DCC can safely expand access to high-value, previously off-limits datasets through governed synthetic alternatives.
- Researchers and startups get realistic data with documented fitness-for-purpose; regulators and stewards get measurable privacy and compliance assurances.
- Synthetic datasets become first-class citizens in the catalog, with provenance, policy, and validation metrics, supporting responsible, equitable AI innovation across the Commonwealth.

Familiarity with AI Fairness Tools

EY brings deep, practical expertise applying AI fairness toolkits and methods across the full model lifecycle (data selection, feature engineering, training, evaluation, deployment, and monitoring), so that equity and accountability are engineered into the DCC by default. Our approach aligns with NIST AI RMF 1.0 and leading responsible AI practices and is implemented with a combination of open-source toolkits, commercial governance platforms, and policy-as-code in the platform.

Toolkits and frameworks we use in practice

- IBM AI Fairness 360 (AIF360): Comprehensive metrics and mitigations for classification and regression; supports pre-, in-, and post-processing techniques (e.g., reweighing, adversarial debiasing, calibrated equalized odds).
- Microsoft Fairlearn: Model comparison across subgroups, threshold tuning subject to fairness constraints, and dashboarding for stakeholder review.
- Google What-If Tool and TensorFlow Model Analysis (TFMA): Slice-based evaluation to visualize performance disparities and conduct counterfactual "what-if" inspections without code.
- Interpretability for fairness diagnostics: SHAP, LIME, and Explainable Boosting Machines (EBMs) to identify proxy features and subgroup-specific drivers of disparate outcomes; DiCE for counterfactual explanations to support remediation.
- Adversarial Robustness Toolbox (ART): Robustness and security testing that complements fairness (e.g., ensuring mitigation strategies don't introduce new vulnerabilities).
- Data auditing and profiling: Facets/Great Expectations/dbt tests to assess representation, label quality, and potential measurement bias in training data.
- LLM and generative model fairness/harms: Evaluation using community datasets and probes and toxicity/harm tests to assess stereotyping, exclusion, and harmful content.

Fairness metrics we operationalize

- Group fairness: Demographic parity, equal opportunity, equalized odds, predictive parity; calibrated AUROC/PR across protected subgroups.
- Individual/counterfactual fairness: Invariance of predictions under counterfactual changes to protected attributes (where causal assumptions are appropriate).
- Error analysis: False positive/negative rates, calibration error, and threshold sensitivity by subgroup to surface disparate impact.
- Distributional checks: Representativeness and coverage (tail behavior, intersectional slices) to avoid masking harms to smaller populations.

Mitigation techniques we employ

- Pre-processing: Reweighing, disparate impact remover, synthetic oversampling with constraints, label de-biasing and relabeling workflows.
- In-processing: Fairness-constrained tuning (regularizers/penalties), adversarial debiasing, monotonic/causal constraints to reduce proxy effects.
- Post-processing: Threshold adjustments, calibrated equalized odds, reject option classification; policy-aware human-in-the-loop overrides where required.
- Causal and policy-aware approaches: Structural causal models for counterfactual fairness when domain knowledge supports causal assumptions; documentation of trade-offs among accuracy, fairness, privacy, and explainability.

Data-level fairness and documentation

- Dataset audits: Coverage of protected and intersectional groups; label noise and measurement bias assessment; proxy/feature leakage detection.
- Synthetic data fairness: Measure whether synthesis preserves or distorts subgroup relationships; evaluate downstream fairness metrics on models trained with synthetic data.
- Transparency artifacts: Datasheets/Data Cards for datasets and Model Cards for models, with explicit fairness testing results, limitations, and intended use.

Integration into platform delivery and MLOps

- Fairness-by-design gates: CI/CD checks that run AIF360/Fairlearn slices on every promoted model; fail conditions and exception paths documented in runbooks.
- Dashboards and reviews: Interactive fairness dashboards embedded in research workspaces; standardized review packs for governance boards and stewards.
- Continuous monitoring: Scheduled slice-based evaluations in production for drift in subgroup performance; alerting and rollback/canary policies when fairness SLOs are breached.
- Policy-as-code: Attribute-based access control (ABAC) and automated enforcement of use restrictions; linkage of fairness test results to model registry approvals.

Ecosystem and alliances

- IBM watsonx.governance integration: Centralized lineage, documentation, fairness metrics capture, and approval workflows for regulated use cases.
- Snowflake integration: Storage of evaluation artifacts and subgroup metadata as governance tags; reproducible fairness evaluations close to data.

How this benefits the DCC

- Consistent, auditable fairness evaluations: Standard test batteries and model cards enables every model and dataset in the DCC is assessed and documented in a comparable way.
- Rapid, responsible experimentation: Researchers gain ready-to-use notebooks and pipelines for fairness metrics and mitigations, reducing the friction to do the right thing.
- Governance-ready outputs: Fairness evidence is machine-readable in the catalog and tied to promotion gates, enabling transparent, equitable AI practices across agencies, academia, and industry.

Illustrative outcomes we have delivered

- Public benefits eligibility: Reduced disparity in false negative rates across demographic groups via pre-processing reweighing and calibrated equalized odds, with minimal impact on accuracy and fully documented trade-offs.
- Healthcare triage: Improved equalized odds across age and race slices by combining causal feature constraints with post-processing, coupled with SHAP-based clinician review to maintain trust.
- NLP moderation for public forums: Decreased toxicity bias against identity terms while maintaining overall precision/recall, validated with RealToxicityPrompts and targeted subgroup tests.

Our familiarity with AI fairness tools is not simply theoretical, we embed these tools into day-to-day workflows, governance processes, and platform automation so that equitable outcomes are measurable, enforced, and sustained throughout the DCC lifecycle.

EY's Alliance Ecosystem

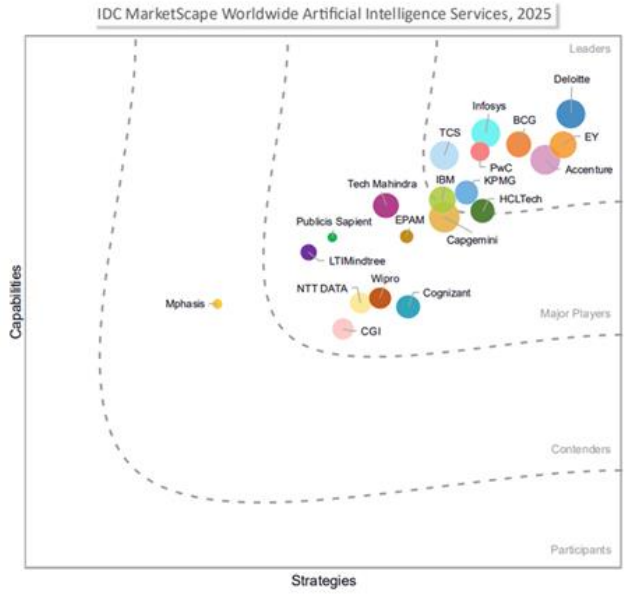
The pace of change has never been faster, the need for resiliency has never been more critical, and embracing transformation has never been more vital.

At EY, we know that these challenges are difficult for a single organization to deal with on its own. That's why EY teams work with an ecosystem of partners, including global strategic alliances, to provide the right technology, capabilities and insights to help clients create long-term value and transform at speed and scale.

For more information, see: https://www.ey.com/en_us/alliances

The strength of our alliances can be seen in the award-winning work EY has done for its clients, and through the recognition of that work in alliance partner awards:

- In 2025 EY was recognized as Snowflake Financial Services Data Cloud Services Partner of the Year for the third consecutive year and Snowflake Media & Entertainment Data Cloud Services Partner of the Year.
- In 2025 EY was recognized as an IBM Partner Plus Global Award Winner for AI for Business for groundbreaking AI compliance solutions.
- In 2024 EY was awarded the prestigious IBM Global Partner of the Year Award for AI for Business for EY's innovative AI platform, powered by Watsonx.
- In 2025 EY was recognized as the #2 global AI Strategy Leader in IDC's MarketScape Worldwide Artificial Intelligence Services Rankings



IDC's leadership rankings | Credit: IDC

The select alliance details below represent a fraction of EY's overall alliance ecosystem, but speak to the strength of EY's offerings in data, cloud services, infrastructure, and AI. These alliance details are meant only to share the story of our alliance ecosystem, and importantly underscore the importance of leaning-in to technology advancement shoulder to shoulder with industry so that we can bring the best solutions to our clients.

Critically important to this effort is our alliance with Snowflake, one of the only commercial technology platforms mentioned in the AI Hub DCC solution RFP materials. Our partnership with Snowflake is detailed below, and in addition in Appendix D where you will find a letter referencing EY's position in the Snowflake Partner Network.

Snowflake

The EY–Snowflake Alliance offers an approach that links business, technology and enterprise elements, enabling clients to more quickly and effectively benefit from the Snowflake Data Cloud. Leveraging our combined experience in data and analytics, cloud operations, industry and adjacent technology ecosystems, we help solve difficult challenges and offer comprehensive data-driven transformations and outcomes.

Alliance-at-a-Glance

As a Snowflake Elite partner, EY teams are empowered to build leading-class Snowflake solutions and provide unparalleled levels of service to our clients. We have over 3,500 Data and AI professionals, and over 2,000 global delivery services data specialists. Within our Data team, more than 545 practitioners are SnowPro certified.

End-to-End Services	Knowledge + Technology	Preferential Support

EY and Snowflake provide clients with a one-stop shop for end-to-end Snowflake offerings.

As our clients look to solve their data challenges, Snowflake technology and EY business use cases drive prioritized execution.

EY SnowPro practitioners have access to advanced Snowflake training and certifications, including:

- Access to latest product roadmaps and prototyping
- Support and access to Snowflake subject-matter resources

Snowflake provides preferential technical support and extends investments to EY engagements.

- Covered by the Master Reciprocal Sub Contract Agreement
- Access to Snowflake Partner Services Fund (PSF)
- Access to solution investment dollars

IBM

Organizations today face growing pressure to modernize legacy systems, transform supply chains, reduce IT costs and adopt AI with speed and precision. The EY–IBM alliance helps address these challenges by combining IBM's advanced technologies with the business experience of EY teams to deliver scalable, secure solutions tailored to each industry.

Whether it's deploying hybrid cloud infrastructure, applying AI across operations or streamlining IT environments, the focus is on helping organizations move faster, work smarter and adapt with confidence. Together, EY teams and IBM provide the tools and insight to drive transformation and support long-term performance.

IBM Platinum Partner

EY and IBM teams combine business ingenuity with AI and hybrid cloud technology to solve the world's most complex business, environmental and social challenges. Our long-term value creation spans sustainability, productivity and interoperability, orchestrating throughout best-in-class IBM software solutions with other technologies that form the backbone of our clients' businesses, like Microsoft Azure Public Cloud or SAP. Our service delivery model comprises both Managed Services and implementation, the latter frequently in teaming with IBM Consulting. The combination of EY as a leading business integrator and IBM Consulting as a renowned systems integrator consistently proves to be a compelling and differentiated proposition to our clients. Together, EY and IBM teams are building a better, more sustainable working world enriched by artificial intelligence (AI) and enabled by hybrid IT.

EY teams have 3,000+ highly skilled practitioners in IBM technologies across the globe, with more than 850 IBM training badges and certifications. This extensive delivery network is supported by a dedicated IBM TechHub@EY, which serves as a skills and engagement enablement center that promotes continuous training, formal certifications and improved client outcomes.

Alliance-at-a-Glance

Alliance Benefits	Alliance Solutions	Delivery Capabilities
▪ More protected client investments and future-proofed businesses ▪ Accelerated growth for ESG ambitions resulting in meaningful, value-led outcomes ▪ AI-infused ways of working while championing the pillars of AI ethics	▪ ESG reporting and decarbonization (aaS) ▪ Enterprise Asset Management ▪ FinOps and cloud economics ▪ AI ops and governance ▪ Quantum Strategies and Roadmaps	▪ 3,000+ IBM skilled EY practitioners, including 850+ IBM training badges and certifications ▪ AI everywhere and unleashing generative AI (GenAI) capability while assuring ethical AI usage including explainability, transparency and privacy; ongoing strategy and use case development of Quantum-based solutions

Microsoft

Every day, throughout the world, businesses, governments, and capital markets rely on EY business ingenuity and the power of Microsoft technology to solve the most challenging global issues. The EY-Microsoft Alliance designs and helps deliver transformative cloud solutions powered by business ingenuity that helps create long-term value. The EY organization and Microsoft bring a compelling formula to spark the potential of the cloud and unlock the power of data. We solve clients' most challenging issues by blending trusted industry knowledge with innovative cloud technology. The strategic relationship draws on decades of success developing visionary solutions that provide lasting value. Together, we empower organizations to create exceptional experiences that help the world work better and achieve more.

Our strategic relationship draws on decades of success developing visionary solutions that provide lasting value, including:

- Sector-led solutions and platforms that bring new business models to bear and accelerated innovation with trust embedded.
- Client proven, repeatable platform solutions that reimagine enterprise functions to be agile, resilient and better equipped to respond to market disruption.
- Data-driven digital transformations that reimagining our client's businesses through custom, human-centered cloud solutions powered by Microsoft technologies.

The EY–Microsoft Alliance offerings generate long-term value by unlocking the power of data and combining business ingenuity, industry experience and intelligent technology. The EY organization has the track record, efficient delivery services, and human capital resources to generate solutions across the Microsoft Cloud. Together, the Alliance helps solve the more complex demands and requirements of the modern business landscape

Alliance-at-a-Glance

Alliance Benefits	Alliance Solutions	Delivery Capabilities
▪ Scale data science across their enterprise ▪ Derive business value from their data ▪ Empower data teams to simplify and democratize data and AI	▪ Advanced enterprise strategy execution ▪ Customer-centric experience transformation ▪ Unified digital operations transformation ▪ Workforce-friendly experience transformation	▪ Providing business-driven consulting in data architecture, data science, engineering, AI, and machine learning. ▪ Industry-specific solutions including Demand Forecasting, Inventory Tuning and Data Fabric supported by Databricks.

ServiceNow

How do you deliver exceptional human experiences that inform, connect and empower? Today's pace of life demands instant transactions, easy access to information and seamless user experiences. Successful organizations quickly address inefficiencies that impact front- and back-office enterprise functions and yield poor experiences for partners, suppliers and customers. The EY organization and ServiceNow unlock value through workflow automation so business works smoother, employees work smarter and customers don't have to work at all.

Alliance-at-a-Glance

The EY–ServiceNow Alliance combines ingenuity with a leading digital workflow platform, following a human-centered and experience-driven approach to innovation. Leveraging the agility of the Now platform and incorporating the power of AI and automation, we can help modernize and link workflows to deliver intuitive and frictionless experiences across your organization.

Alliance Benefits

- Digitize and automate siloed processes to significantly improve experiences for your private and public sector employees, customers and third parties.
- Allow employees to focus on meaningful work by using in-platform AI and analytics to automate routine tasks, predict and avoid incidents, and surface insights.
- Risk and resilience can become manageable real-time, as ServiceNow connects the business, security and IT with fully integrated risk management.
- Control your supply chain, manufacturing operations and OT security on a single platform.
- ServiceNow brings 85% of the Fortune 500® smarter, faster and better ways to work, making them fit for a fast-changing world.

Delivery Capabilities

- Over 2,800 ServiceNow certifications across the globe.
- Average CSAT score of 4.4/5 on ServiceNow deployments.
- Ability to spin up minimum viable products in less than five days.
- A 50/50 onshore/offshore delivery model allows us to be price-competitive with GSIs, boutiques and outsourcing firms.
- Our ServiceNow services collaborate with solution teams, service lines and sector leaders to bring an end-to-end transformation.
- Harvesting assets and accelerators from past projects allows us to deliver rapid time to value for our customers — in many cases, we go live in under three months.

Databricks

With a growing number of companies looking to champion data in their business strategies, there is a need for a platform that caters to data transformation, data science and analytics. Addressing that need, the EY–Databricks Alliance provides such companies with a platform to help with the rapid delivery of value at scale, complemented with technology consulting and training. Furthermore, the Alliance helps companies drive their business strategies to success through massive-scale data engineering, collaborative data science, full-lifecycle machine learning and business analytics.

Alliance-at-a-Glance

The EY–Databricks Alliance helps clients transform their organizations by enabling new digital capabilities that unlock value from their data. This Alliance also sells services and technology together and support each other to deliver value for key client relationships.

The EY organization has a significant number of skilled consultants on Databricks technology to help clients drive better business outcomes. In addition, the EY organization participates in the development of sector solutions and accelerators that further leverage Databricks technology and allow for faster go-to-market capabilities.

Alliance Benefits	Alliance Solutions	Delivery Capabilities
▪ Scale data science across their enterprise ▪ Derive business value from their data ▪ Empower data teams to simplify and democratize data and AI	▪ Machine Learning Operations (MLOps) ▪ Data Mesh ▪ Predictive Maintenance ▪ Secure Research Enclaves ▪ Demand Forecasting and Inventory Tuning ▪ Intelligent Financial Forecasting and Scenario Modeling	▪ Providing business-driven consulting in data architecture, data science, engineering, AI, and machine learning. ▪ Industry-specific solutions including Demand Forecasting, Inventory Tuning and Data Fabric supported by Databricks.

Approach

Overview

Our track record of success within the Commonwealth is demonstrative of the fact that we understand the unique needs and challenges faced by government entities in Massachusetts. The proposed approach that follows takes into consideration our extensive experience working alongside state agencies to transform them through innovation programs. Where other large firms may see opportunity to over-engineer and extract value for themselves, and boutique firms lack the diverse and nuanced experience necessary to deliver a well architected, extensible, maintainable, and secure end-to-end solution, EY offers a thoughtful, measured approach where we place value to Massachusetts Technology Collaborative first.

Outlined in the DCC Requirements.PDF provided with the RFP package was a sample phased timeline. EY's proposal focuses on starting with the "Initiation & Design" portion of that timeline. By carving this out as a Phase 0, Massachusetts Technology Collaborative will realize multiple tangible benefits and avoid numerous risks.

Presented in the Q&A was a consistent message indicating vendors should propose recommendations and appropriate solutions. Appropriateness is a key consideration that necessitates our recommended approach.

Absent of detailed user journeys, functional requirements, defined workflows, baseline technical, infrastructure, and security requirements, it is not yet feasible to align architectural decisions, tooling, and implementation estimates as appropriate.

For this reason, we propose an engagement under which we will collaborate with Massachusetts Technology Collaborative to document and refine these requirements and inform evidence-based solution decisions and an accurate, predictable implementation roadmap.

Failure to adopt this approach risks inappropriate architecture and technology choices that either underserve functionality or provide expensive, unused features that overcomplicate future platform extensibility and maintainability. Either way the result is extensive re-work, streams of change orders, runaway costs, and at worst, program failure.

In comparison, adopting our proposed approach enables the up-front collaboration necessary to create a predictable, cost-effective, implementation roadmap which is derived from an appropriately designed solution architecture aligned with fully defined requirements and Massachusetts Technology Collaborative's value drivers.

Working with EY under this approach significantly de-risks delivery. Further, our Agile methodology for iterative development that follows during implementation phases enables accelerated delivery putting capabilities in the hands of end-users early, consistently, and often. The flexibility afforded by this approach means Massachusetts Technology Collaborative can continually monitor user-feedback to direct project investments towards the most meaningful outcomes.

Detailed Project Approach and Plan

Discovery & Implementation Readiness (Phase 0) - Overview

Building a robust platform requires collaborative teaming with the right vendor. There are multiple potential solution approaches which would achieve the objectives of the platform. As such, a measured approach is to begin with a Discovery phase in which Massachusetts Technology Park and EY will collaborate over a 16-week period. This Discovery and Implementation Readiness planning Phase 0 will enable deep discussions on goals, features, architecture, technology selection and alignment across stakeholders. This will drive success in future build phases. This collaborative approach will lead to more effective outcomes and budget expectations. Providing a rough estimate rooted in multiple assumptions would not ultimately align with goals and expectations in a cost effective manner. The figure below illustrates EY's hybrid-Agile delivery methodology. This approach has proven the most successful in delivering transformation initiatives such as these, especially when working with public sector agencies such as Massachusetts Technology Collaborative. We have highlighted "We are here" to indicate where our combined team currently sits along the process methodology map.

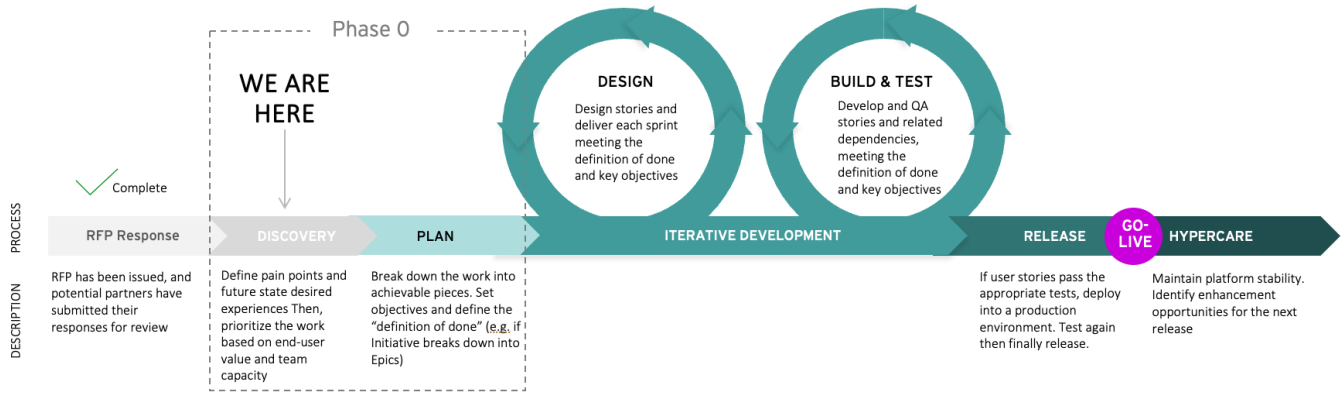

Phase 0
WE ARE HERE
DESIGN
Design stories and deliver each sprint meeting the definition of done and key objectives
BUILD & TEST
Develop and QA stories and related dependencies, meeting the definition of done and key objectives
Complete
PROCESS
RFP Response
DISCOVERY
PLAN
ITERATIVE DEVELOPMENT
RELEASE
GO-LIVE
HYPERCARE
DESCRIPTION
RFP has been issued, and potential partners have submitted their responses for review
Define pain points and future state desired experiences Then, prioritize the work based on end-user value and team capacity
Break down the work into achievable pieces. Set objectives and define the "definition of done" (e.g. if Initiative breaks down into Epics)
If user stories pass the appropriate tests, deploy into a production environment. Test again then finally release.
Maintain platform stability. Identify enhancement opportunities for the next release

Figure 1: *Iterative delivery methodology*

During Discovery we propose a collection of key areas that will lead us to implementation-readiness and surface a cost effective delivery solution and roadmap. The areas include:

- **Requirements analysis deep dive sessions -** Guides the future state roadmap and architecture we collaborate with key stakeholders to define detailed future state aspirations and distill down to capabilities required to enable the future state.
- **Developing the future state architecture to deliver expected results -** Based on defined detailed requirements, we collaborate on technology selection and data requirements needed to build a future state digital platform in a secure, scalable and cost effective manner.
- **Gathering requirements for future state user experience -** Based on defined detailed requirements, we collaborate on user personas, workflows and user journeys so that we will be able to meet the goals of delivering a responsive, multilingual, accessible, and fair platform.
- **Building a detailed roadmap to guide the implementation priorities –** We define a sequence of build activities required to realize future state vision and timeline horizons, the conceptual architecture that incorporates all functional future state capabilities, and data and integration priorities.
- **Alignment sessions –** We conduct review sessions with program leadership to discuss findings and align on goals, recommendations, and feasibility.
- **Socialization and Approval Sessions -** Conduct review sessions to confirm approach directionality, roadmap priorities & feasibility along with build readiness.

This collaboration will lead to well-defined implementation readiness and set up the program for successful delivery. Furthermore, this approach enables unlocking strategic value across the solution ecosystem enabling all stakeholders involved to achieve their goals and manage costs by leveraging the right tools for the job with a preference toward open-source solutions for added community support and cost savings. Discovery and implementation readiness planning will enable:

- **Clear strategic alignment** - on the architecture, technology, and user experience will enable the development of a solution and roadmap that captures the highest value for MassTech DCC and its user community.
- **Relevant future state architecture** - that is scalable across the ecosystem in alignment with stated requirements.
- **Actionable understanding of steps to realize future state** - in a tangible and time-bound way considering capacity, dependency and milestone requirements that meet business and technology aspirations.
- **A balanced solution roadmap -** aligned with prioritized requirements to effectively manage implementation costs.

We have assembled a purpose-built team for this effort that brings together deep understanding of the MassTech stakeholder landscape, knowledge of open source and AI ecosystem, and experience collaborating with large teams, state technology and vendor agencies locally and globally.

Below in **Figure 2**, we summarize our proposed high-level plan for the Discovery & Implementation Readiness Phase 0.

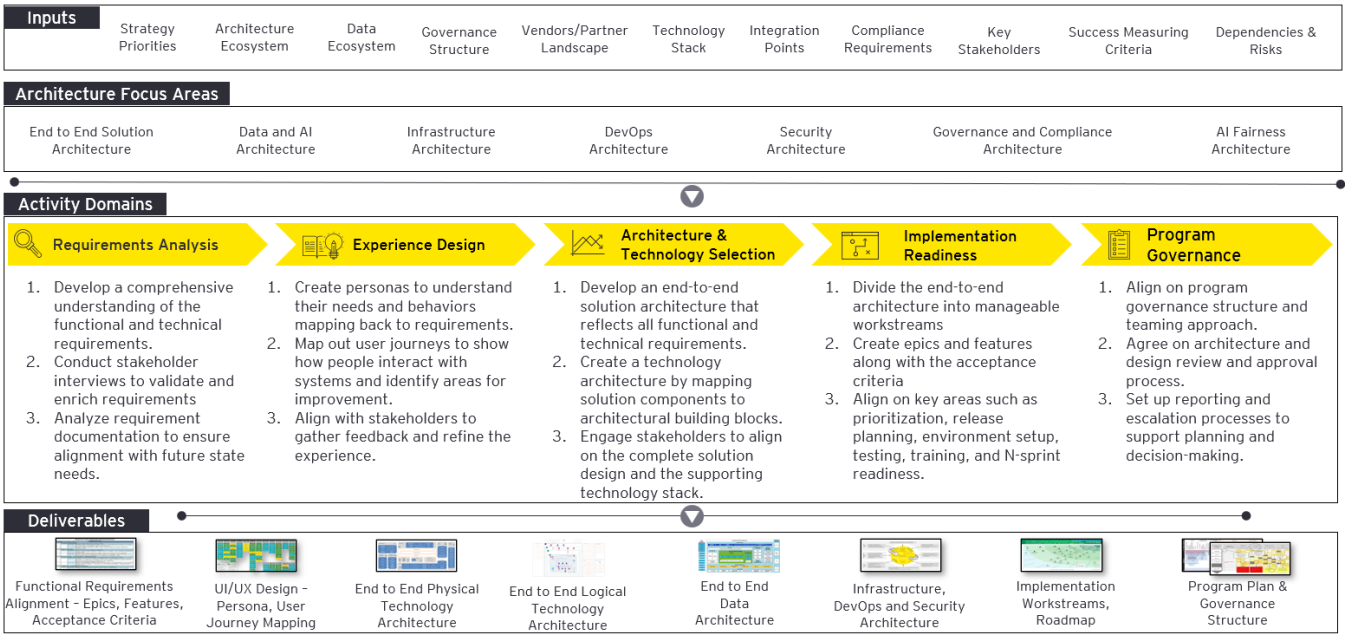

Inputs
Strategy Priorities
Architecture Ecosystem
Data Ecosystem
Governance Structure
Vendors/Partner Landscape
Technology Stack
Integration Points
Compliance Requirements
Key Stakeholders
Success Measuring Criteria
Dependencies & Risks
Architecture Focus Areas
End to End Solution Architecture
Data and AI Architecture
Infrastructure Architecture
DevOps Architecture
Security Architecture
Governance and Compliance Architecture
AI Fairness Architecture
Activity Domains
Requirements Analysis
1. Develop a comprehensive understanding of the functional and technical requirements.
2. Conduct stakeholder interviews to validate and enrich requirements
3. Analyze requirement documentation to ensure alignment with future state needs.
Experience Design
1. Create personas to understand their needs and behaviors mapping back to requirements.
2. Map out user journeys to show how people interact with systems and identify areas for improvement.
3. Align with stakeholders to gather feedback and refine the experience.
Architecture & Technology Selection
1. Develop an end-to-end solution architecture that reflects all functional and technical requirements.
2. Create a technology architecture by mapping solution components to architectural building blocks.
3. Engage stakeholders to align on the complete solution design and the supporting technology stack.
Implementation Readiness
1. Divide the end-to-end architecture into manageable workstreams
2. Create epics and features along with the acceptance criteria
3. Align on key areas such as prioritization, release planning, environment setup, testing, training, and N-sprint readiness.
Program Governance
1. Align on program governance structure and teaming approach.
2. Agree on architecture and design review and approval process.
3. Set up reporting and escalation processes to support planning and decision-making.
Deliverables
Functional Requirements Alignment - Epics, Features, Acceptance Criteria
UI/UX Design - Persona, User Journey Mapping
End to End Physical Technology Architecture
End to End Logical Technology Architecture
End to End Data Architecture
Infrastructure, DevOps and Security Architecture
Implementation Workstreams, Roadmap
Program Plan & Governance Structure

Figure 2: *High level Discovery and Implementation Readiness Phase 0 Approach*

Activities

Our proposed Discovery phase 0 includes a comprehensive plan of action such that time spent together is of high value and leads to aligned actionable next steps, assets, and outcomes. EY has prepared the following plan of activities that work to enable a successful Discovery phase. Thoughtful detailed planning ahead of and throughout Discovery will enable successful implementation phases.

As outlined in each of the sections that follow, we've highlighted key activities for each iteration across a set of key functional domains. These activities have been carefully considered to allow our combined team to recognize challenges and opportunities, glean collaborative insights from key stakeholders and EY's experiences, and iterate on our approach to make sure it is both relevant and impactful. By thoroughly understanding the target audiences, their needs, and the current landscape, our solution and implementation approach can be tailored to stakeholder expectations.

The activities across the workstreams outlined below combine synergistically to cover program and project management across each of the following requirement areas outlined in the DCC Rquirements.PDF

- 4.0 Target Users
    - AI/ML developers and researchers
    - Academic (K-12, higher education) institutions and research labs
    - State government departments and civic tech groups
    - Startup and enterprise AI Teams
    - Nonprofits working with Massachusetts open data
- 5.0 External Datasets and Data Capabilities
    - Webpage
    - Dataset Directory
    - Dataset Detail Pages
    - Community Dataset Submission
    - Admin Dashboard
    - Accessibility, Performance, Security, Responsiveness, & Extensibility
- 6.0 Website Functionality for Hosted Data and Access
    - Data Hosting & Access
        - Searchable Datasets
        - Metadata standards
        - Programmatic APIs
    - User Experience Enhancements
        - Dataset explorer with filters
        - Preview functionality
        - Access requests
- 7.0 Commons Credit System

- 8.0 Data Catalog & Metadata Management System
    - Metadata Management
    - Dataset Discovery & Search
    - Access Control & Governance
    - Dataset Publishing & Versioning
    - Collaboration & Community Features
    - Data Preview & Exploration
- 9.0 Dataset Curation Plan
- 10.0 Synthetic Data Generation Capability
- 11.0 AI Fairness & Bias Capability
- 12.0 Governance
- 13.0 Security
- 14.0 Governance Policies and Ethical Guidelines

Activities for each of the seven (7) workstream pillars:

Workstream 1: Program & Project Management (PMO)

	Week	Key Activities
Kickoff	Week 1	• Establish kickoff, workshop and collaboration meetings cadence • Establish regular leadership reporting check ins • Establish weekly status reports • Engage stakeholders and schedule workshops • Review activity plan, timeline, and expectations with each team
Sprint 1	Week 2	• Facilitate functional and non-functional requirements capture in PM tools to enable later backlog generation • Facilitate cross team collaboration sessions and attendance
	Week 3	
Sprint 2	Week 4	• Define Ways of Working • Define RACI
	Week 5	
Sprint 3	Week 6	• Define program operating model • PM tools technology selection and facilitate tools availability
	Week 7	
Sprint 4	Week 8	• Facilitate early Agile training and program setup operations PM tools technology selection and facilitate tools availability • Oversee and validate backlog generation
	Week 9	
Sprint 5	Week 10	• Oversee and validate backlog generation • Define training requirements
	Week 11	
Sprint 6	Week 12	• Facilitate backlog prioritization sessions
	Week 13	
Sprint 7	Week 14	• Facilitate Pre-PI planning workshops • Lead the prioritized roadmap creation and socialization
	Week 15	
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Workstream 2: Experience Design

	Week	Key Activities

Kickoff	Week 1	• *Kickoff, onboarding & stakeholder identification* • *Stakeholder collaboration sessions*
Sprint 1	Week 2 Week 3	• *Establish catalog of UI screens for the solution* • *Establish user persona mapping to understand requirements for:* ○ *Homepage, Dataset Directory, Dataset Detail Pages, Public Datasets, Synthetic Data Tools, Massachusetts Public Datasets*
Sprint 2	Week 4 Week 5	• *Establish user persona mapping to understand requirements for:* ○ *Community Dataset Submission, Admin Dashboard, Additional Requirements* • *Design workshop*
Sprint 3	Week 6 Week 7	• *Establish user persona mapping to understand requirements for:* ○ *Data Hosting & Access, User Experience Enhancements, Credit Solution, Using MGHPCC compute resources* • *Define user journey maps and workflows* • *Being wireframing capabilities capturing initial low fidelity mock-ups for review*
Sprint 4	Week 8 Week 9	• *UI/UX implementation approach and technology selection*
Sprint 5	Week 10 Week 11	• *Define initial backlog of tasks to define high fidelity design and UI components*
Sprint 6	Week 12 Week 13	• *Prioritized backlog by phase*
Sprint 7	Week 14 Week 15	• *Review sessions and design workshop*
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Workstream 3: Security

	Week	Key Activities
Kickoff	Week 1	• Stakeholder interview to understand all system and facility security requirements • Evaluate existing security capabilities & tooling to identify reuse option
Sprint 1	Week 2 Week 3	• Identify data, cloud, data center and code security needs both functional and non-functional • Capture requirements documentation
Sprint 2	Week 4 Week 5	• Define security monitoring and alerting requirements • Define security checklist for developers

Sprint 3	Week 6 Week 7	• Define encryption and role based access control (RBAC) requirements • Collaborate with DevOps on CI/CD security integrations
Sprint 4	Week 8 Week 9	• Define data protection and classification requirements.
Sprint 5	Week 10 Week 11	• Security tooling technology selection • Define initial implementation backlog • Document procedure to adhere to industry security best practices
Sprint 6	Week 12 Week 13	• Prioritized backlog by phase
Sprint 7	Week 14 Week 15	• Review sessions
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Workstream 4: Cloud & Infrastructure

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding, stakeholder identification • Stakeholder collaboration sessions • Socialize & gain alignment on scope and objectives
Sprint 1	Week 2 Week 3	• Validate functional requirements
Sprint 2	Week 4 Week 5	• Non-functional requirements planning (performance, HA/DR, compliance) • Identify cloud provider & on-prem infrastructure requirements
Sprint 3	Week 6 Week 7	• Review implementation methodology (provisioning, deployments, hybrid practices) • Draft high-level architecture diagrams (cloud landing zone, hybrid connectivity)
Sprint 4	Week 8 Week 9	• Technology/tool validation for services and interoperability
Sprint 5	Week 10 Week 11	• Calculate running hosting & usage costs (based on tools and expected load) • Define initial implementation backlog
Sprint 6	Week 12 Week 13	• Prioritize backlog items by phase (Seed → Grow → Bloom) • Review sessions with stakeholders
Sprint 7	Week 14 Week 15	• Establish governance framework (decision-making, reporting cadence) • Define compliance/security baselines (logging, monitoring, encryption, retention policies) • Draft risk register and mitigation strategies

Discovery Closeout	Week 16	• Socialize and gain solution approvals • Define implementation readiness next steps • Develop training/knowledge transfer plan • Prepare runbooks for Day-1/Day-2 operations • Change management & communication planning

Workstream 5: Web Portal & Credit System

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions
Sprint 1	Week 2 Week 3	• UX and user workflow • Detailed functional requirements for: ○ *Homepage, Dataset Directory, Dataset Detail Pages, Public Datasets, Synthetic Data Tools, Massachusetts Public Datasets*
Sprint 2	Week 4 Week 5	• Non-functional requirements planning • Detailed functional requirements for: ○ *Community Dataset Submission, Admin Dashboard, Additional Requirements* • Detail feature capabilities to understand full scope needs
Sprint 3	Week 6 Week 7	• Evaluate options to align implementation costs with value drivers • Begin solution architecture definition capturing all system and tools that will encompass the system
Sprint 4	Week 8 Week 9	• Detailed functional requirements for: ○ *Data Hosting & Access, User Experience Enhancements, Credit Solution, Using MGHPCC compute resources* • Experience design collaboration to size UX implementation & requirements • Portal architecture definition & review continued
Sprint 5	Week 10 Week 11	• Define QA testing practices • Define initial backlog • Technology selection enabling efficient development with a lean towards open source technologies as accelerators and cost savers
Sprint 6	Week 12 Week 13	• Prioritized backlog by phase
Sprint 7	Week 14 Week 15	• Collaborate with DevOps to define local developer SDLC to support efficient onboarding and development • Review sessions
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Workstream 6: Data & Data Services

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions
Sprint 1	Week 2 Week 3	• Interviews for details discovery of initial data capabilities ○ Data hosting and Access, Data Curation Plan, Metadata management, Dataset Discovery and Search, Access Control & Governance
Sprint 2	Week 4 Week 5	• Interviews for details discovery of initial data capabilities ○ Dataset Publishing & Versioning, Collaboration & Community Features, Data Previews & Exploration, Synthetic Data Generation, AI Models, Commons Credits

Sprint 3	Week 6
	Week 7
		• Detailed data requirements
		  ○ Data Assets priorities
		  ○ Data Services priorities
		  ○ Non-functional data requirements planning
		• Governance Requirements

Sprint 4	Week 8
	Week 9
		• Design Data Capabilities and Services for enablement and integration
		  ○ Define Op Model/Governance for data assets
		  ○ Data Publishing
		  ○ Data Maintenance
		  ○ Data Subscriptions
		  ○ Etc.
		• AI Fairness Requirements

Sprint 5	Week 10 Week 11	• Data implementation approach and technology selection • Define initial backlog of tasks to define high fidelity data design and data services components
Sprint 6	Week 12 Week 13	• Prioritized backlog by phase
Sprint 7	Week 14 Week 15	• Review sessions and design workshop
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Workstream 7: DevOps

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions
Sprint 1	Week 2 Week 3	• Evaluate tool preferences

Sprint 2	Week 4 Week 5	• Identify data, cloud, data center and code deployment needs both functional and non-functional • Capture requirements documentation
Sprint 3	Week 6 Week 7	• Identify data, cloud, data center and code observability needs both functional and non-functional •
Sprint 4	Week 8 Week 9	• Define CI/CD pipeline requirements for each deployment target and code feedback loops • Technology selection of specific tools, services and facility interoperability
Sprint 5	Week 10 Week 11	• Calculate running hosting & usage costs based on tools and traffic load expectations • Define local developer SDLC workflow • Define initial implementation backlog
Sprint 6	Week 12 Week 13	• Prioritized backlog by phase
Sprint 7	Week 14 Week 15	• Review sessions
Discovery Closeout	Week 16	• Socialize and gain solution approvals • Implementation readiness next steps

Deliverables

Below are high-level details for each of the deliverables depicted in **Figure 2:** High level Discovery and Implementation Readiness Phase 0 Approach. These include:

- Functional Requirements
- UI/UX Design
- End to End Physical Solution Architecture
- End to End Logical Technology Architecture
- End to End Data Architecture
- Infrastructure, DevOps and Security Architecture
- Implementation Workstream Roadmap
- Program Plan & Governance Structure

Functional Requirements

Product Backlog

EY will follow an Agile process that will begin in the Discovery phase. The functional requirements elicited during discovery will take the shape of an Agile product backlog organized by Epics, Features, and User Stories. User Stories will outline the details of a requirement and include high-level Acceptance Criteria (AC). The AC will be further refined during the Implementation phases. The Product Backlog will be created in Massachusetts Technology Collaborative's preferred DevOps tool, such as Azure DevOps or JIRA.

Key Focus: Requirements gathering, AC definition, and backlog organization

Deliverable: Product backlog

UI/UX Design

EY's Experience Design group are a critical part of the requirements gathering process. They excel at translating business requirements into intuitive user interface designs which delight users and elevate user experience.

Key Focus: Translation of requirements into Persona definitions, user Journey Maps, and resulting Wireframes and Visual Designs.

Personas

Based on our qualitative and quantitative user research findings, the EY team develops composite descriptions of representative users (personas) – each representing a unique set of needs, objectives, likely behaviors and attitudes. Unlike use cases, a persona accounts for a typical set of tasks and preferences, extending beyond a role or an actor to help us formulate tasks that they typically need to complete. These personas will be leveraged throughout the design and delivery phases to help our teams and key stakeholders build empathy with the end user, tell stories about the end user and to stress test our recommendations.

What a persona tells us:

- Who are my users?
- What motivates my user to use (or not use) a product?
- What behavioral patterns does my ideal user exhibit?
- What challenges does my user face?

Why we build personas:

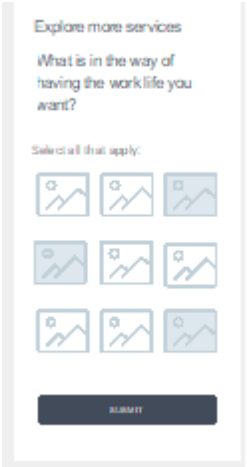
- To build empathy
- To reduce design bias
- To create a shared understanding
- To help identify & prioritize features by user type
- To guide your design decisions from start to finish.
- To provide living documents that can be regularly revisited as you learn more about your users.

Journey Maps

Our researchers and designers work together to create journey maps based on data collected from end users. Journey maps identify key users and their journey through completing specific critical tasks. Journey maps identify user interactions with your product and team, their pain points, areas of delight and potential opportunities. A journey map is a design artifact that is used to focus conversations on solutions and bring empathy to the users of your product. Later the team creates and consolidates the future state journeys represented by relevant personas to develop an integrated map of key interactions and attitudes along the service delivery lifecycle**.**

Wireframes/Visual Designs

Using the Personas and Journey Maps as informative guides, we will design screen layouts. As shown below (left), we start with low-fidelity wireframes to present iterative options and explore alternatives. After gathering input on these wireframes, we move quickly to high-fidelity mockups with accurate representations and sample content, as shown below (right). These higher fidelity wireframes and mockups illustrate functional requirements will inform development at the user-story level.



Sample Low-Fidelity Wireframe ***Sample High-Fidelity Mock-up***

Figure 3: *Sample Wireframe and High-Fidelity Mock-Up*

Mockups of individual pages or components are usually sufficient to gain consensus with core stakeholders and guide developers. For usability test subjects and stakeholders unfamiliar with the design/development process, before development, we will create "clickable" design prototypes that simulate select functions or screen flows with sample content.

Deliverables: Persona documentation, Journey Map documentation, Wireframes and Visual Designs

End to End Physical Solution Architecture

The AI Hub Data Commons Collaborative plans to span a range of physical components and locations including MGHPCC, Cloud & Third Party solutions therefore exploring and designing the details of where capabilities will physically reside, how these locations communicate and securing them is vital.

This refers to the actual hardware and infrastructure that supports a system. It can include, for example, but is not limited to:

- Servers (on-premises or cloud-based)
- Network devices (routers, switches, firewalls)
- Storage systems
- Data centers
- End-user devices (desktops, mobile devices, REST API interfaces, developer tools, etc.)

Key Focus: Where components are deployed, how they are connected, and how data physically flows through the system.

Technology Selection

Our technology selection process is designed to validate the chosen solutions align with the overall architecture strategy and business objectives. In performing technology selection there are several key factors in the EY selection framework that include:

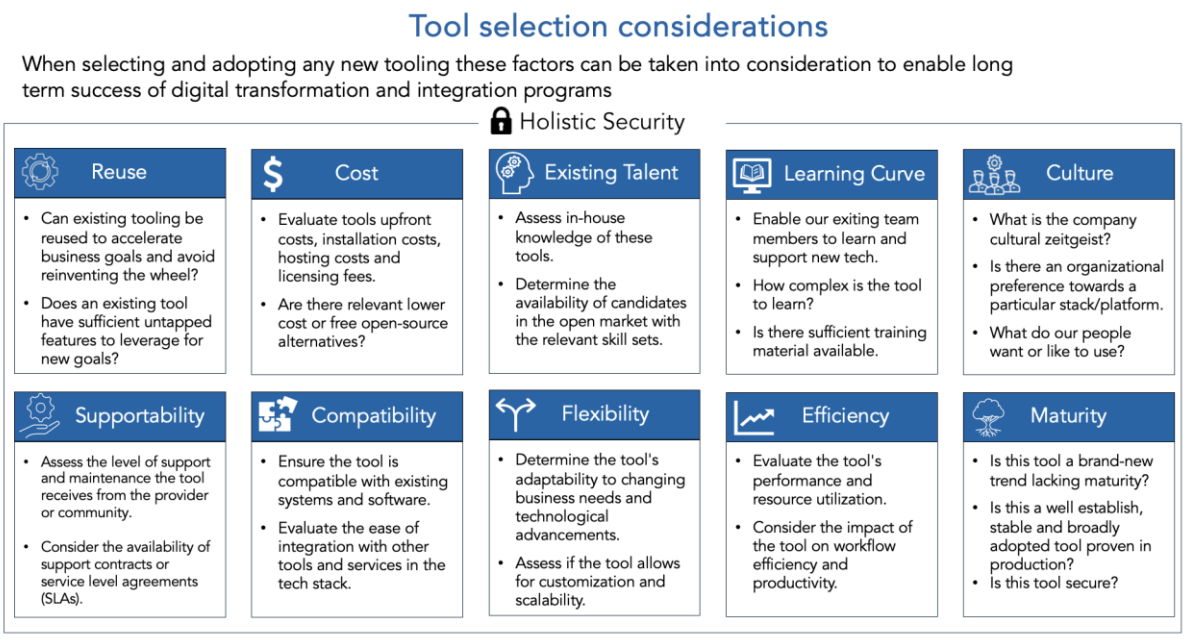

Tool selection considerations
When selecting and adopting any new tooling these factors can be taken into consideration to enable long term success of digital transformation and integration programs
Holistic Security
Reuse
Can existing tooling be reused to accelerate business goals and avoid reinventing the wheel?
Does an existing tool have sufficient untapped features to leverage for new goals?
Cost
Evaluate tools upfront costs, installation costs, hosting costs and licensing fees.
Are there relevant lower cost or free open-source alternatives?
Existing Talent
Assess in-house knowledge of these tools.
Determine the availability of candidates in the open market with the relevant skill sets.
Learning Curve
Enable our exiting team members to learn and support new tech.
How complex is the tool to learn?
Is there sufficient training material available.
Culture
What is the company cultural zeitgeist?
Is there an organizational preference towards a particular stack/platform.
What do our people want or like to use?
Supportability
Assess the level of support and maintenance the tool receives from the provider or community.
Consider the availability of support contracts or service level agreements (SLAs).
Compatibility
Ensure the tool is compatible with existing systems and software.
Evaluate the ease of integration with other tools and services in the tech stack.
Flexibility
Determine the tool's adaptability to changing business needs and technological advancements.
Assess if the tool allows for customization and scalability.
Efficiency
Evaluate the tool's performance and resource utilization.
Consider the impact of the tool on workflow efficiency and productivity.
Maturity
Is this tool a brand-new trend lacking maturity?
Is this a well establish, stable and broadly adopted tool proven in production?
Is this tool secure?

Figure 4: *Technology tool selection consideration dimensions*

Deliverables: Technology Architecture narrative document including diagram(s) showing how web servers, application servers, and databases are distributed across different geographic locations. See the Quality and Clarity of Proposed Technical Architecture section for a sample of our early thoughts on proposed architecture for the Data Commons Collaborative.

End to End Logical Technology Architecture

The AI Hub Data Commons Collaborative plans to span a range of logical components as well including a web portal, API's, data and many others therefore exploring and designing the details of functional relationships, communications contracts, capabilities ownership and securing them is vital.

This represents the abstract design of the system; how components interact conceptually, regardless of where or how they are physically deployed. It can include, for example, but is not limited to:

- Modules and layers (e.g., presentation, business logic, data access)
- Services and APIs
- Data flow and control flow
- Security models and access control

Key Focus: Functional relationships, responsibilities, and interactions between components.

Deliverables: Technology Architecture narrative document including diagram(s) showing the flow of interaction between the end users and physical system components.

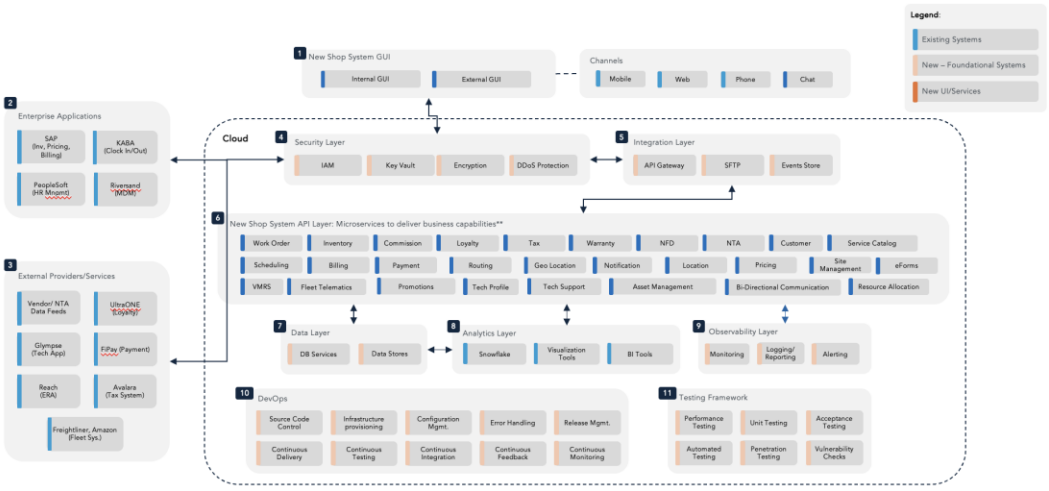


Figure 5: *Sample logical architecture of the systems and tools that will encompass the solution*

End to End Data Architecture

Our end-to-end data architecture provides a governed, scalable, and secure path for data from diverse sources through ingestion, curation, and enrichment into consumption layers for analytics and reporting, with continuous monitoring of data quality, performance, and cost. It organizes the platform into clear zones and cross-cutting capabilities so teams can build reliable products with predictable SLAs.

Data flows from sources through governed ingestion and a raw landing zone, into curated enterprise datasets, and onward to tuned serving layers and an internal data marketplace for governed discovery and access. Cross-cutting metadata, lineage, orchestration, security, observability, and FinOps span every zone to foster reliability, compliance, and performance.

- Sources ⟶ Ingestion: Operational DBs, SaaS, files, events/IoT enter via batch/stream/CDC with contracts and schema validation.
- Landing (Bronze): Immutable, source-aligned storage; classification/tags; basic data quality checks and metadata capture.
- Curation (Silver): Standardize, deduplicate, enrich; conform to canonical models; apply reference data; enforce DQ SLAs; incremental transforms.
- Serving (Gold): Publish marts, semantic models, APIs, and feature sets; aggregates/materialized views for fast, governed reporting.
- Consumption: Structured dashboards/reports and analytics sandboxes for ad-hoc, data science, and ML via a governed semantic layer.

- Metadata & Lineage: Central catalog with business/technical metadata; end-to-end lineage and impact analysis; schema registry for events.
- Orchestration: DAGs and event triggers manage dependencies, retries, backfills, promotions, and SLAs.
- Security & Privacy: IAM/RBAC/ABAC, encryption, masking/tokenization, policy-as-code, auditing across all zones.
- Observability & Performance: Pipeline SLIs/SLOs, DQ scorecards, query latency/scan metrics, cost attribution/guardrails; alerts and runbooks.

Key Focus:

Scale ingestion (batch/stream/CDC) with data contracts, embed data quality in curation, and enable metadata-driven, secure self-service via a semantic layer and an internal data marketplace, backed by metadata management and end-to-end lineage. Performance, FinOps, and operability (SLAs, alerts, auto-recovery) are built in by design.

Deliverables:

Reference Data Architecture Document: Includes diagrams illustrating interactions between data source systems, data management capabilities, and data consumption. Metadata Framework: Specifies critical metadata requirements for tagging and certifications. Data Operating Model: Highlights various personas and their interactions with data.

End to End Reference Data Architecture (Illustrative):

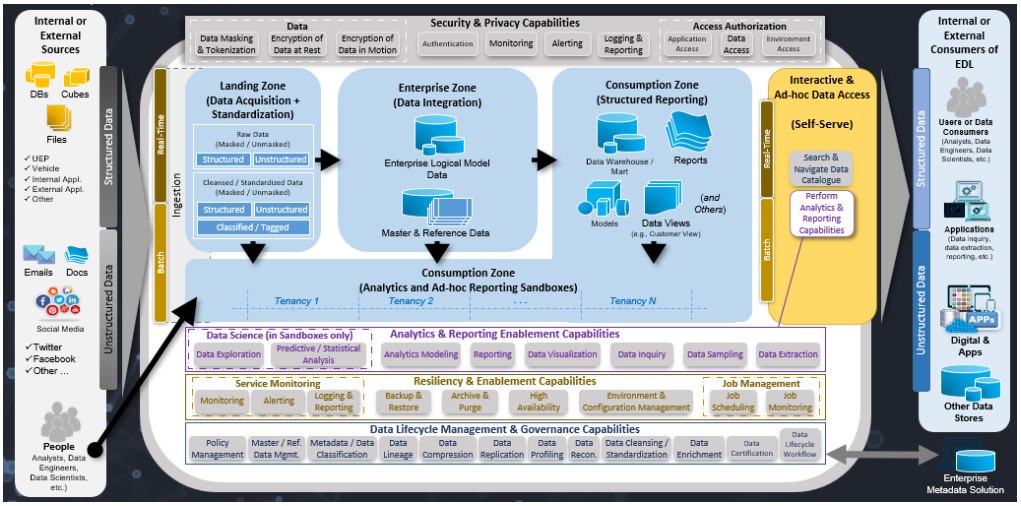


Figure 6: *High level End to End Architecture*

Infrastructure, DevOps, and Security Architecture

Infrastructure Architecture

Infrastructure Architecture defines the foundational hardware and cloud resources required to support the Data Commons Collaborative platform. It includes specifications for hybrid cloud deployment (on-premises and cloud-based servers), networking components, storage systems, and data center configurations. The architecture enables scalable compute resources (CPU/GPU), secure networking, workload isolation, and high-throughput data processing, with automated provisioning and environment parity across development, test, and production environments using Infrastructure as Code.

Key Focus:

Deliver a comprehensive Hybrid Cloud Architecture Blueprint and supporting documentation that defines infrastructure requirements and provisioning approach for the Data Commons Collaborative platform.

Deliverables:

- **Hybrid Cloud Architecture Blueprint: High-level diagrams showing cloud landing zones, hybrid connectivity, and networking topology.**
- **Infrastructure Requirements Document: Specifications for compute, storage, networking, HA/DR, and compliance considerations.**
- **Provisioning Approach Overview: Conceptual methodology for Infrastructure-as-Code and automated environment setup (no actual scripts).**

DevOps Architecture

The DevOps architecture outlines the automated processes for building, testing, deploying, and monitoring platform services and data pipelines. It incorporates CI/CD pipelines for code and container management, security scanning, policy enforcement, and continuous integration of new features. The approach includes standardized workflows for system integration, process updates, thorough checks, monitoring, and observability, ensuring reliable and repeatable deployments and rapid feedback loops for development teams. The diagrams on following pages show a possible outline of this approach.

Key Focus:

Provide a CI/CD Pipeline Design Blueprint and governance framework outlining deployment strategy, toolchain integration, and monitoring concepts for reliable and scalable operations.

Deliverables:

- **CI/CD Pipeline Design Document** - Blueprint of pipeline stages for build, test, deploy, and monitoring, including security and policy checkpoints.
- **Toolchain Integration Plan** - Documented validation of selected tools and interoperability strategy (e.g., GitHub Actions, Azure DevOps, Terraform).

Security Architecture

The Security architecture outlines a layered approach in support of comprehensive protection against threats while maintaining compliance and facilitating incident response by integrating Identity and Access Management (IAM) to control user access, employing encryption to protect data at rest and in transit, utilizing data discovery and classification tools to identify and categorize sensitive information, and implementing logging and monitoring systems to track and analyze activities.

Key Focus:

Provide a Secure Data Management Blueprint and governance framework outlining deployment strategy, technology integration, and monitoring capabilities for securely managing sensitive data.

Deliverables:

- **Secure Data Governance Framework -** A holistic modern approach to governing and protecting data assets throughout the data lifecycle, inclusive is data discovery, data classification, data encryption, access management and monitoring capabilities
- **Identity and Access Management IAM reference architecture –** A reference architecture to enable logging and audit for user management role activities, conduct access reviews, implement analytics for behavior deviations, and automate access certification for timely reviews and approvals

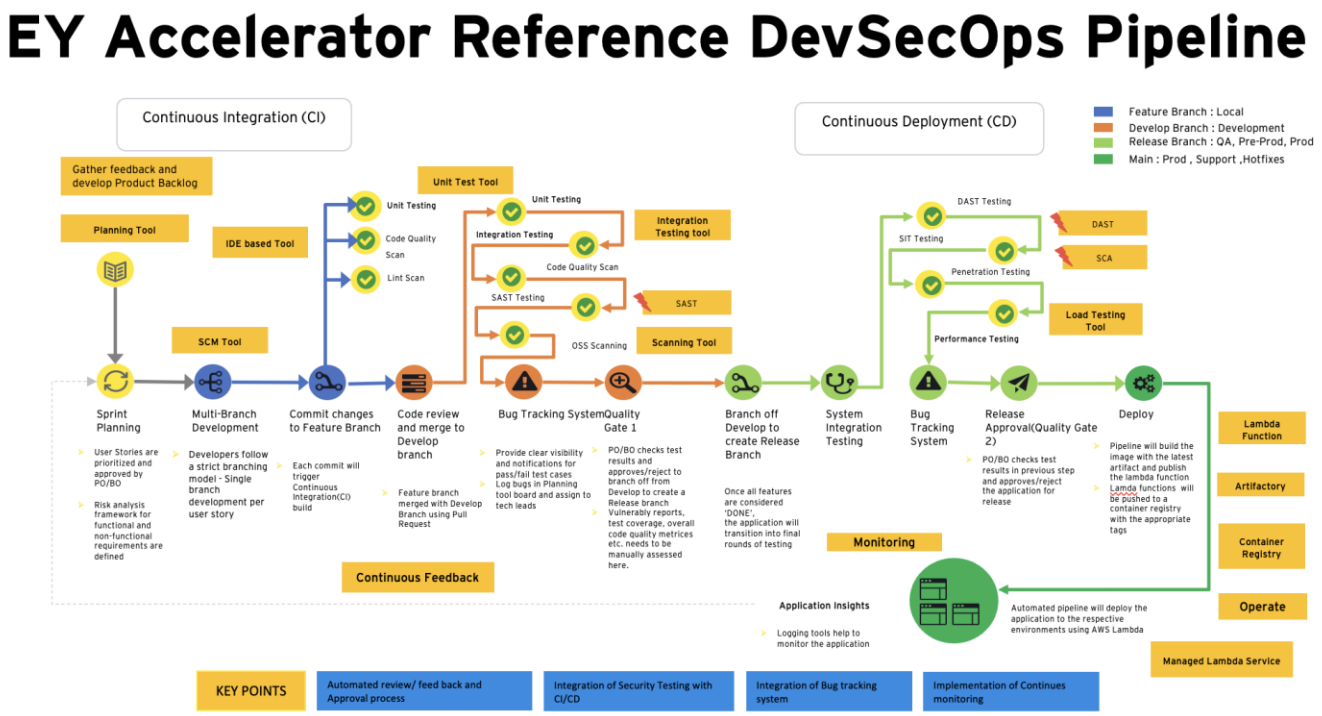


Figure 7: *Sample DevSecOps Pipeline Flow*

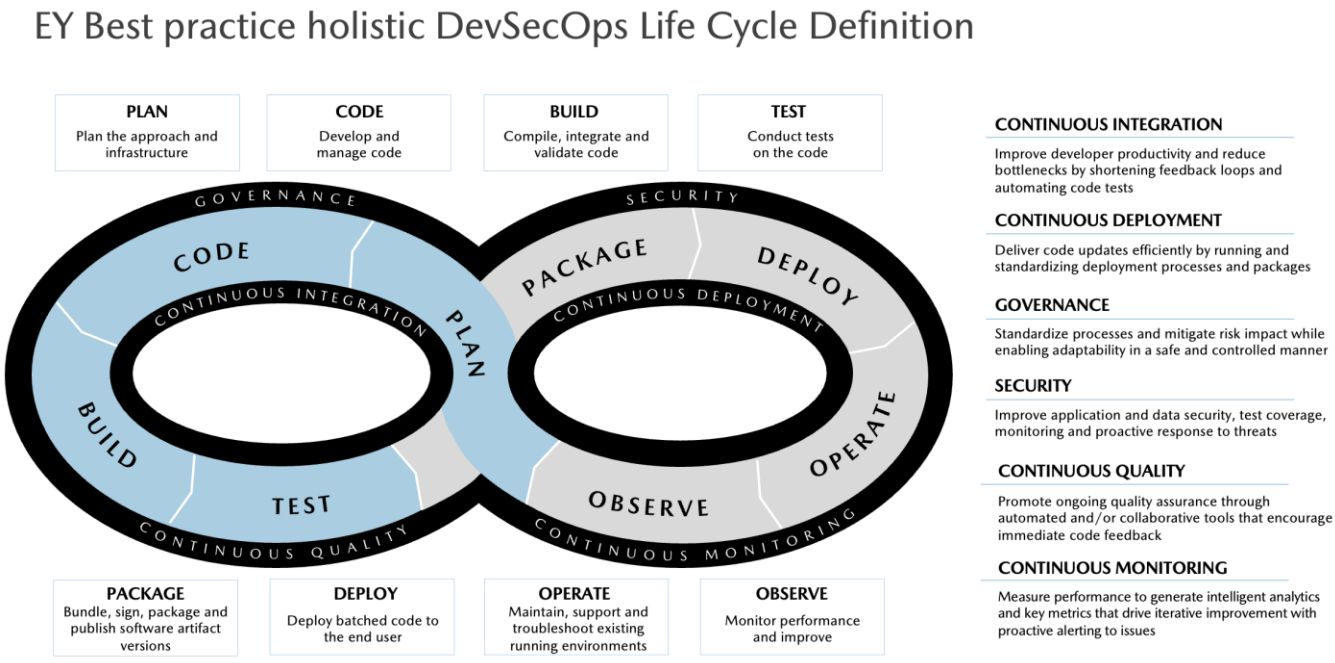

EY Best practice holistic DevSecOps Life Cycle Definition
PLAN
Plan the approach and infrastructure
CODE
Develop and manage code
BUILD
Compile, integrate and validate code
TEST
Conduct tests on the code
GOVERNANCE
CODE
CONTINUOUS INTEGRATION
PLAN
BUILD
TEST
CONTINUOUS QUALITY
SECURITY
PACKAGE
DEPLOY
CONTINUOUS DEPLOYMENT
OPERATE
OBSERVE
CONTINUOUS MONITORING
PACKAGE
Bundle, sign, package and publish software artifact versions
DEPLOY
Deploy batched code to the end user
OPERATE
Maintain, support and troubleshoot existing running environments
OBSERVE
Monitor performance and improve
CONTINUOUS INTEGRATION
Improve developer productivity and reduce bottlenecks by shortening feedback loops and automating code tests
CONTINUOUS DEPLOYMENT
Deliver code updates efficiently by running and standardizing deployment processes and packages
GOVERNANCE
Standardize processes and mitigate risk impact while enabling adaptability in a safe and controlled manner
SECURITY
Improve application and data security, test coverage, monitoring and proactive response to threats
CONTINUOUS QUALITY
Promote ongoing quality assurance through automated and/or collaborative tools that encourage immediate code feedback
CONTINUOUS MONITORING
Measure performance to generate intelligent analytics and key metrics that drive iterative improvement with proactive alerting to issues

Figure 8: *Sample DevSecOps Lifecycle*

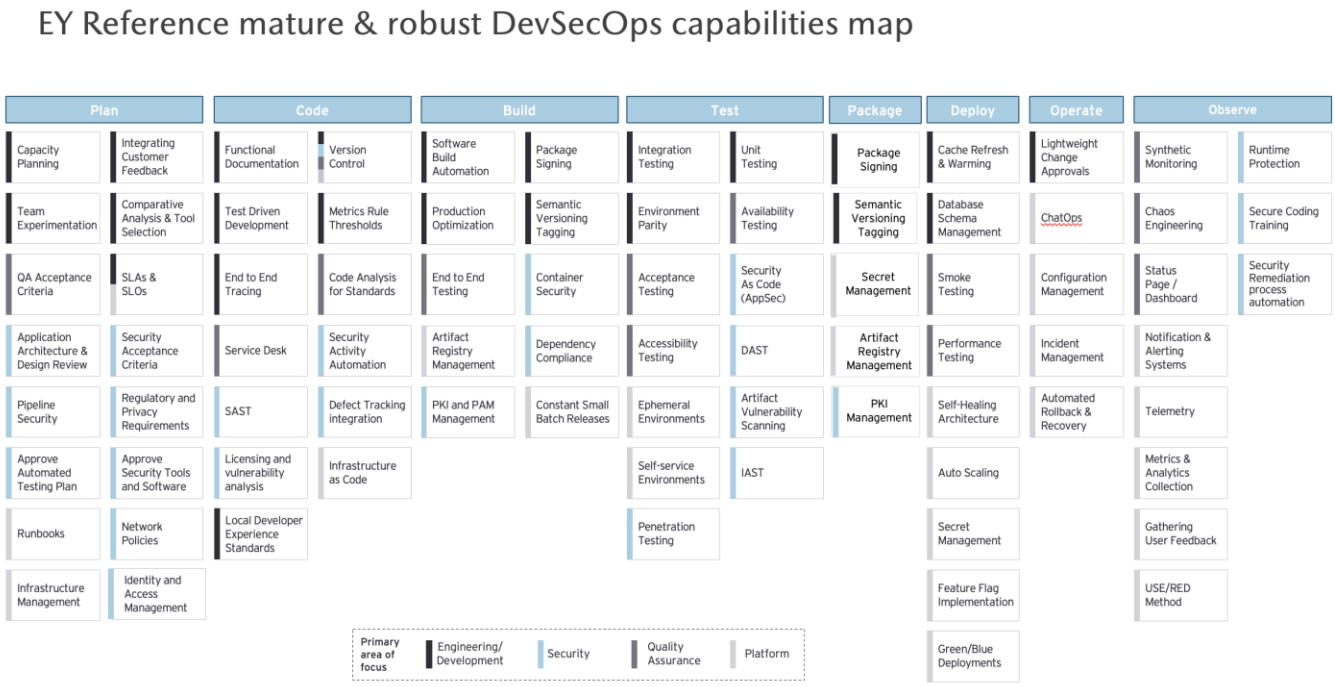


Figure 9: *DevSecOps capabilities matrix*

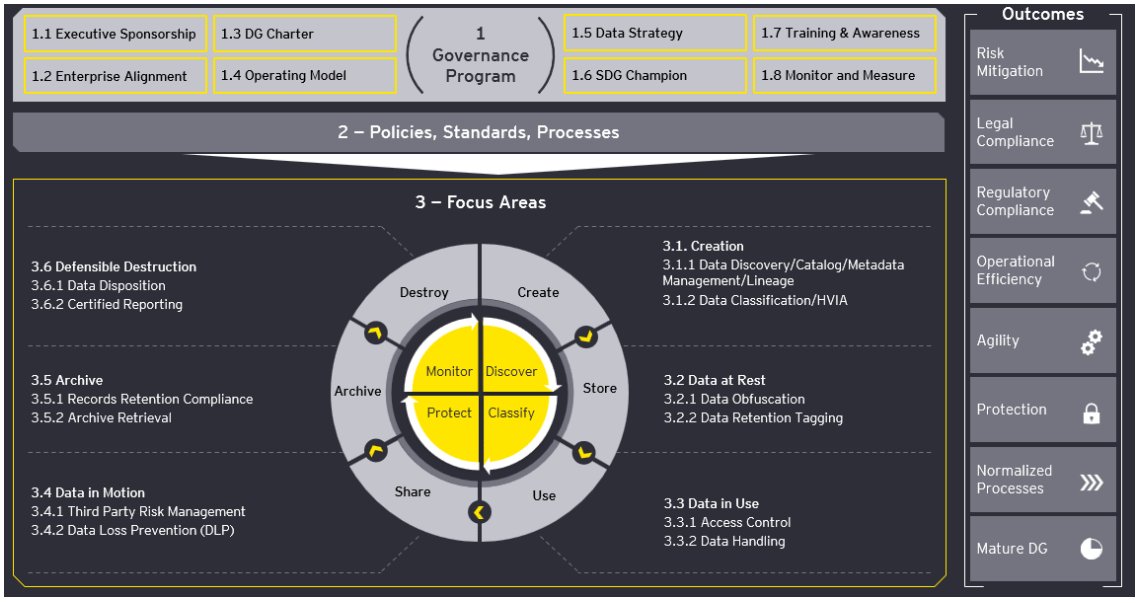


Figure 10: *Secure Data Governance Framework*

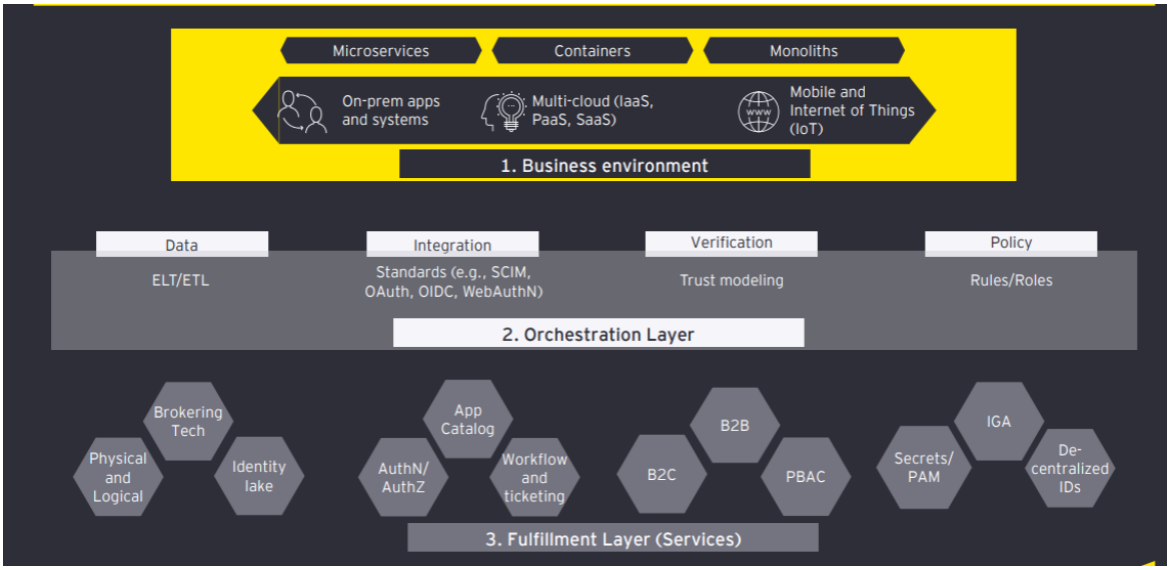


Figure 11: *Identity and Access Management IAM reference architecture*

Implementation Workstream Roadmap

Creating a project roadmap provides a clear, visual plan for sequencing activities, aligning stakeholders, and managing dependencies, ensuring the project stays on track and meets its objectives within scope, time, and budget. As a part of Discovery & Implementation Readiness Phase 0 we plan to collaborate around roadmap creation while accounting for key readiness activities that include:

- **Clarity & Alignment –** So that all stakeholders understand the project phases, milestones, and deliverables.
- **Dependency Identification -** Identifies critical paths and interdependencies early to avoid bottlenecks.
- **Implementation Resource Planning -** Helps allocate people, tools, and budget effectively across phases.
- **Risks Identification -** Surfaces potential risks and allows for contingency planning.
- **Progress Tracking -** Provides a baseline for monitoring progress and adjusting timelines as needed.

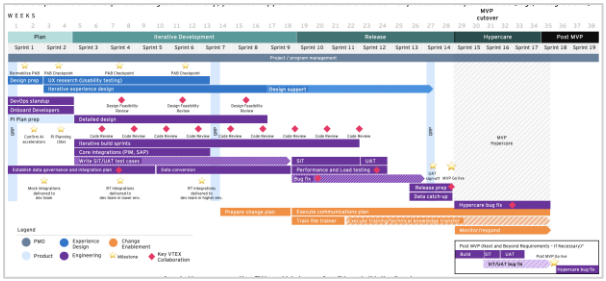
Key Focus:

Activities in planning and developing a roadmap will include:

- **Defining Objectives & Scope -** Confirm what success looks like and what's in/out of scope for each of the phases of the program.
- **Break Down Phases** - Segment work into logical phases based on priorities and when we plan to deliver features to the community.
- **Identify Key Milestones -** Highlight critical checkpoints (e.g., architecture sign-off, MVP release).
- **Map Dependencies -** Document which tasks rely on others to start or finish.
- **Estimate Timelines -** Assign realistic durations based on complexity and resources.
- **Assign Responsibilities -** Clarify ownership for each deliverable and milestone.
- **Incorporate Risk & Contingency -** Build buffers for high-risk activities.

Deliverables:

- **Visual Timeline -** A clear roadmap showing phases, milestones, and dependencies.
- **Narrative document –** Outlines agreement on priorities, timelines, ways of working, and responsibilities.
- **Program performance metrics tracking -** Baseline for tracking progress and reporting status.



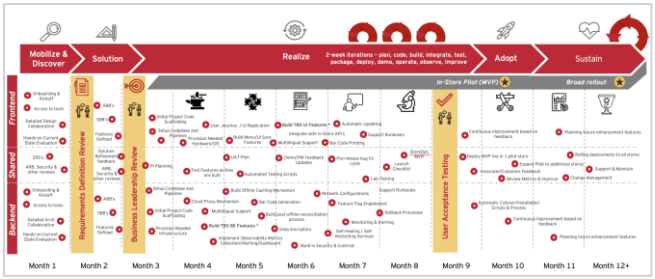


Figure 12: *Sample visual project timeline*

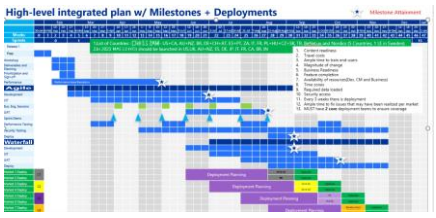
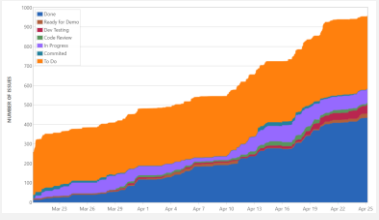
Program Plan & Governance Structure

To support the DCC's objectives and provide robust program oversight, the PMO will establish a comprehensive suite of artifacts, reports, and tooling. These deliverables will empower stakeholders with visibility into project progress, risks, and change management activities, while facilitating efficient communication and accountability across all service areas and workstreams.

Key Focus: Definition of program/project plan and governance structure that will guide Massachusetts Technology Collaborative and EY along this journey.

Deliverables:

- **Integrated Project Plan** - A master plan, maintained in Excel, will capture all project milestones and key dependencies. This plan will serve as the single source of truth for phase planning and execution.
- **Milestone List by Phase** - A detailed list of milestones, organized by project phase, will drive progress tracking and enable proactive risk mitigation.
- **RAID Log** - Risks, Issues, Assumptions, Decisions, and Dependencies will be tracked in a central log to facilitate timely escalation and resolution.
- **RASCI Chart** - A matrix clarifying Responsible, Accountable, Support, Consult, and Inform roles for all major deliverables providing clarity in ownership and accountability.
- **Weekly Status Reports** - Regular reporting will keep stakeholders informed of current status, accomplishments, and emergent risks.
- **Master Table of Meetings** - This artifact will list meeting attendees, cadence, and expected outcomes to coordinate collaboration and governance.
- **Organizational Chart** - Visual charts will clarify team structures, reporting lines, and interdependencies.
- **Onboarding/Offboarding Documentation** - Comprehensive guides will facilitate seamless transitions for team members joining or leaving the project.

Illustrative *integrated project plan* (Excel)	
Illustrative Agile workstream *cumulative flow diagram* (JIRA)	

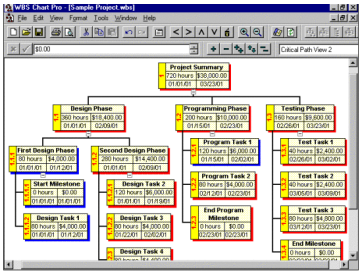
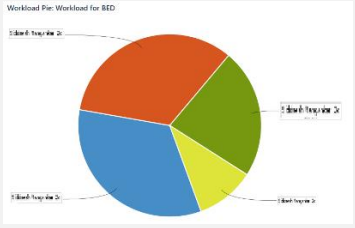
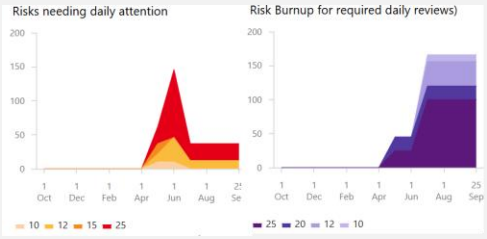
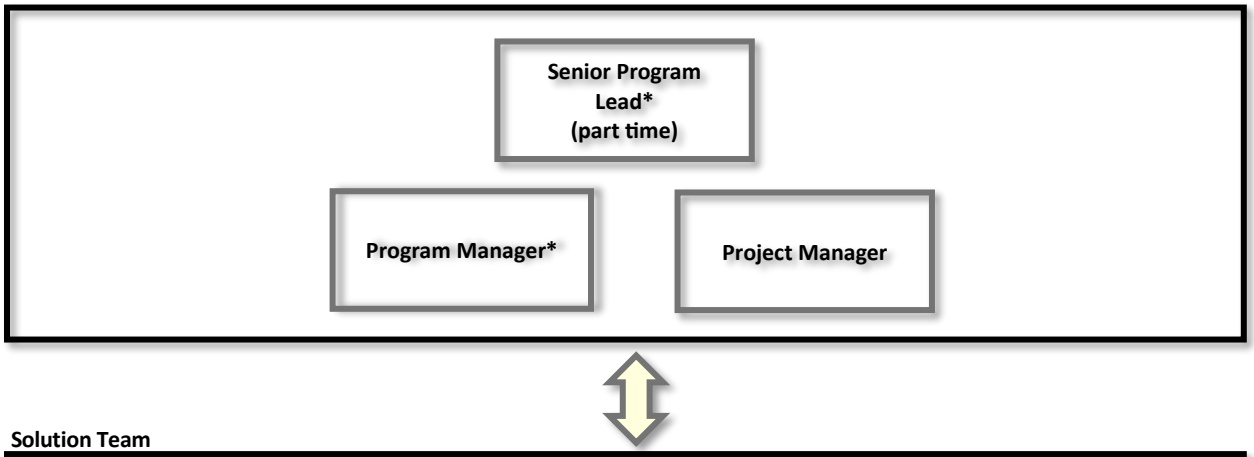
Illustrative Waterfall / blended workstream Work Breakdown Structure (WBS Chart Pro)	
Illustrative Waterfall / blended workstream *remaining-workload pie-chart* (JIRA)	
Illustrative *Roll-Out Plan and checklist* (Excel)	
Illustrative *Risk burndown and burnup charts* (Azure DevOps)	

Figure 13: *Sample project management dashboards*

By deploying these artifacts and leveraging modern tooling and dashboards, the PMO will provide the foundation for transparent governance, continuous improvement, and successful delivery of the DCC's objectives. Each visual asset—integrated project plans, flow diagrams, WBS charts, pie-charts, rollout checklists, and risk visualizations—will be embedded within stakeholder communications and reports, ensuring timely access to actionable information and supporting a culture of collaboration and excellence.

Team Structure

Program Leadership



Solution Team

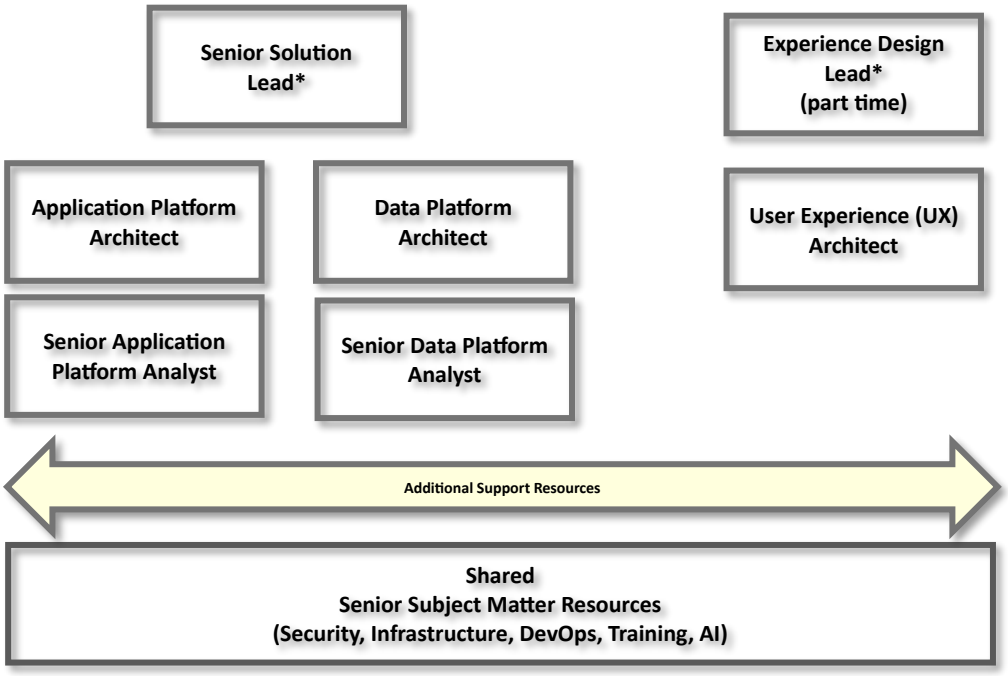


Figure 14: *Proposed Engagement Team Structure*

**Indicates a Key Leadership role. All Key Leadership roles and select additional representative bios can be found in the Resumes & Bios section of this document.*

Roles & Responsibilities

The table below shows the responsibilities for each of the roles defined above.

Role	Responsibilities
Senior Program Lead	• Accountable for project delivery • Manages staffing and administrative aspects • Manages budget and schedule • Escalation point for significant issues • Provides architecture and design oversight
Program Manager	• Works closing with the Senior Program Lead and Project Manager to monitor project progress • Actively works to mitigate risk • Reviews key project materials • Escalation point • Meets with and provides counsel to client sponsors and executives • Manages the overall program

Role	Responsibilities
Project manager	• Day-to-day operational responsibility for project delivery • Manages technical resources • Primary liaison to client • Manages business requirements • Works closely with client project manager to track project progress and direction • Coordinates resources across functional teams • Manages the overall project
Senior Solution Lead	• Provides end-to-end oversight across solution architecture streams • Aligns solution design with strategic priorities and compliance needs • Guides backlog prioritization and technical feasibility • Escalation point for design conflicts across teams • Reviews deliverables for scalability and alignment with success criteria
Application Platform Architect	• Defines application architecture and validates interoperability • Leads platform tool evaluation selection • Collaborates with DevOps on CI/CD pipeline design • Breaks down technical enablers for backlog refinement • Validates application layer meets UX, security, and integration needs
Senior Application Platform Analyst	• Captures detailed application requirements (functional and non-functional) • Documents user stories and acceptance criteria • Defines testing scenarios and validation requirements • Tracks sprint progress against platform deliverables • Analyzes performance, usability, and accessibility requirements
Data Platform Architect	• Defines end-to-end data platform architecture and governance • Validates compliance with FAIR principles • Designs data publishing, versioning, and access control capabilities • Aligns security on encryption and monitoring requirements • Validates scalability for analytics, AI/ML, and reporting
Senior Data Platform Analyst	• Documents detailed data requirements and governance needs • Translates requirements into epics, features, and user stories • Conducts gap analysis for current vs. target data capabilities • Collaborates on dashboards and reporting integration • Validates datasets for accuracy, lineage, and compliance
Senior Experience Design Lead	• Facilitates persona workshops and creates journey maps • Leads design sprints for wireframes and prototypes • Aligns UX with data visualization and reporting needs • Validates accessibility and usability compliance • Collaborates with architects to confirm design feasibility
User Experience Architect	• Defines UX framework and reusable design components • Translates journey maps into UI design patterns • Collaborates with developers for design implementation • Validates UX meets accessibility and security standards • Supports design reviews and backlog validation
Shared Senior Subject Matter Resources	• Provide expertise in security, data governance, DevOps, and compliance • Validate solution approach against DCC federal requirements • Support backlog grooming with best practice guidance • Participate in risk assessments and develop mitigation strategies • Advise on training, change management, and operational readiness

Timeline

Activation of our approach includes a collection of parallel tracks, each following a series of sequential collaborations between EY and Massachusetts Technology Collaborative stakeholders. This methodical approach will enable us to rapidly and effectively discover, solution, and plan for a phased agile implementation increments that derive early value from project investment.

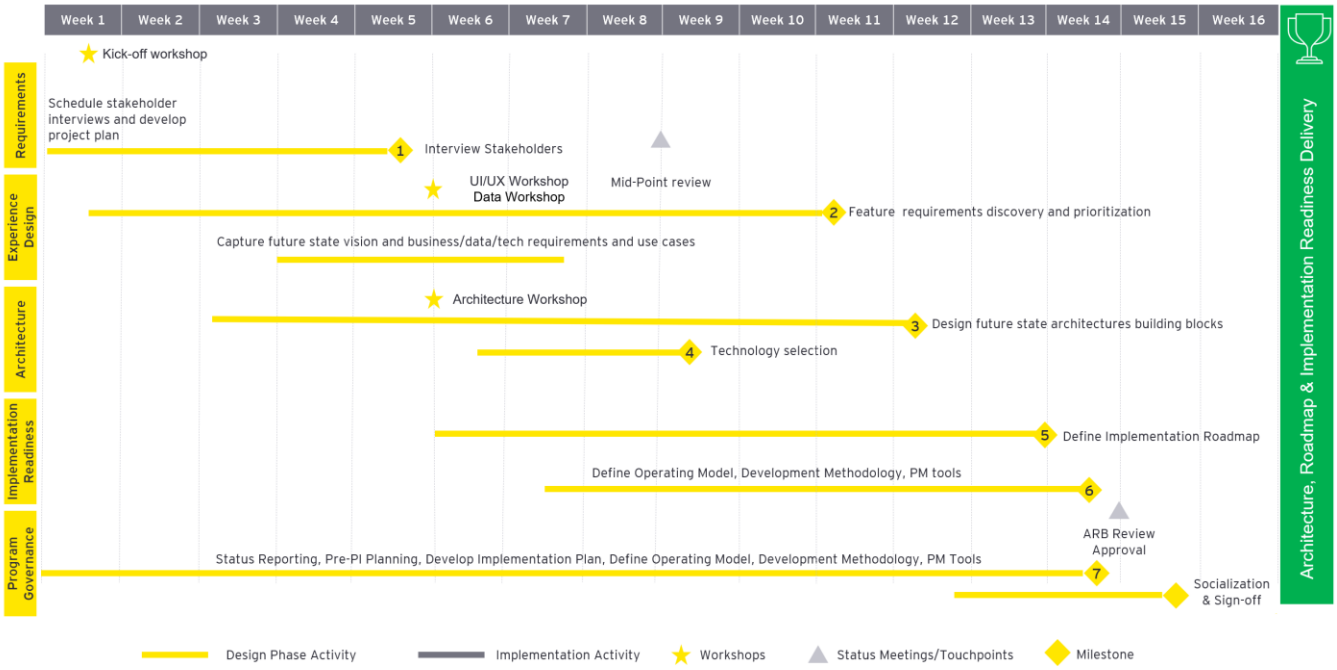


Figure 15: *Proposed timeline of activities and order of operations*

The proposed discovery timeline allows for the stakeholders to collaborate and operate as "One Team" as a whole fostering buy-in, ownership and commitment to our goals for Massachusetts and the broader AI community.

Drawing on our extensive experience managing complex, multi-stakeholder initiatives, we recognize that MassTech's success depends on effective coordination across leaders, vendor personnel, and external organizations. Our approach to project management focuses on tracking progress, identifying interdependencies, aligning every activity with CPI's program goals, delivering the structure and agility to achieve your objectives. EY will begin this effort by leading a Phase 0 kickoff to establish a foundation for success. The kickoff will define strategies around the following core principles to provide structure, accountability, and efficiency throughout the engagement:

- **Strong Governance:** We begin by clearly defining roles and responsibilities such that each stakeholder is accountable for their critical piece of the effort.
- **Goal Alignment:** The Project Team will coordinate with MassTech to establish goals for the project, ensuring they are clearly defined, agreed upon, attainable, and measurable.
- **Project Tracking:** We will collaborate to create a robust Project Plan to identify critical tasks, subtasks, and dependencies. We recognize that MassTech team members have established responsibilities, making it critical to provide a predictable schedule of involvement. To track progress, the team will submit weekly status reports, in a format agreed upon with MassTech, to provide status updates based on all identified tasks and activities included in the Project Plan.
- **Appropriate Scope:** We will collaborate with MassTech to align on Project Scope, defining efforts that are realistic, understood by all, and managed centrally by the EY Program Manager and the core team.
- **Risk Mitigation:** Our program management structure will support clear line of sight to track tasks, identify risks, and develop mitigation plans if the need arises.
- **Outcome-focused:** Our approach will focus on realizing business benefits for MassTech such that we are working towards actionable deliverables.

On Day 1, our team will have the details organized and necessary activities completed – recurring meetings calendared, Project Plan drafted, and the leadership team aligned and engaged – so that we can quickly begin pushing the effort forward. By the end of Phase 0, we will provide implementation readiness assets and begin execution of implementation.

Feasibility of Timeline to Complete Program Scope

EY recognizes the importance of this initiative to the Massachusetts Technology Collaborative and the Commonwealth as a whole. We believe our proposed approach and timeline will accelerate successful delivery. Starting with our proposed 16-week Discovery phase positions our combined teams to effectively meet the outlined program scope, right size requirements, and achieve each of the goals articulated in the RFP through subsequent implementation phases.

Based on our experience with successful delivery of myriad large scale data platform transformation initiatives, and our thorough understanding of the RFP, Data Commons Collaborative Requirements, and provided Q&A, we are confident that this approach is meeting the Massachusetts Technology Collaborative where you currently are and sets us up for shared success.

Below we outline seven (7) key workstreams. We anticipate a weekly commitment of approximately fifteen (15) to (20) hours per week from each of the key stakeholders associated with each workstream. If stakeholders are shared across workstreams the required commitment would be additive.

Our proposed 16-week timeline is in recognition of the fact that personnel at organizations such as yours typically have time constraints driven by their day-to-day responsibilities. 16-weeks aligns with a moderate commitment from Massachusetts Technology Collaborative personnel.

However, with our dedicated resources we are positioned to move more quickly through the Discovery phase on a compressed 12-week timeline, for a lower cost, if Massachusetts Technology Collaborative can dedicate additional time week to week.

We have provided a view of this compressed timeline in Appendix B – Compressed Timeline for review and consideration.

High-Level Technical Architecture Considerations

Based on our current understanding of the solution technical requirements, EY sees 9 key subject areas in the overall architecture which include:

1. **MGHPCC (Massachusetts Green High Performance Computing Center):** Serves as a foundational infrastructure component providing high-performance compute and storage resources for hosting curated datasets, synthetic data generation, and AI model experimentation in a secure, scalable environment.
2. **Cloud Hosting:** The solution leverages a public or hybrid cloud environment to provide scalable, secure, and API-driven infrastructure for hosting curated datasets, synthetic data generation tools, and user-facing services, ensuring elasticity, high availability, and integration with external data sources.
3. **Third-Party Tools:** The solution plans to integrate external open-source and commercial tools such as **CKAN/Dataverse** for dataset cataloging, **SDV/Synthea/Gretel.ai** for synthetic data generation, and **AIF360/Fairlearn** for fairness and bias analysis, enabling modular, standards-based capabilities without building these complex features from scratch.
4. **DevOps Requirements:** Implement a CI/CD pipeline with automated testing, containerized deployments (Docker/Kubernetes), infrastructure-as-code (Terraform), centralized monitoring (Prometheus/Grafana), and security compliance checks to provide scalable, reliable, and auditable delivery of the DCC platform across hybrid cloud and HPC environments.
5. **AI and Data Tools:** The solution will integrate open-source and commercial tools such as **SDV, Synthea, Gretel.ai** for synthetic data generation, **CKAN/Dataverse** for dataset cataloging and metadata management, and **AIF360, Fairlearn, SHAP, LIME** for fairness auditing and explainability, enabling secure data sharing, bias detection, and privacy-preserving AI development within the DCC platform.
6. **Compute Credits:** The solution uses a **Commons Credits System** as a token-based mechanism to incentivize contributions and manage resource usage, allowing users to **earn credits** by uploading datasets or tools and **spend credits** on premium datasets, synthetic data generation, or compute-intensive AI tasks, ensuring fair access and sustainability within the DCC ecosystem.
7. **Developer Tools & APIs:** The solution requires **RESTful APIs, SDKs, and developer-friendly interfaces** for programmatic dataset search, metadata retrieval, synthetic data generation, and Commons Credits management, enabling integration with external systems, Jupyter notebooks, and AI workflows for seamless automation and interoperability.
8. **Secure SDLC Expectations:** The project mandates a security-first development lifecycle with practices such as threat modeling, secure coding standards, automated vulnerability scanning, dependency checks, penetration testing, encryption of data in transit and at rest, RBAC enforcement, and continuous compliance monitoring integrated into CI/CD pipelines so that the platform meets state and federal security and privacy requirements.
9. **Project & Program Management Needs:** The solution requires structured governance with clear milestones, agile delivery practices, stakeholder alignment, risk and compliance tracking, and robust reporting dashboards to facilitate phased implementation, timely decision-making, and adherence to security, ethical, and regulatory requirements throughout the program lifecycle.

Based on the needs of the AI Hub Data Common Collaborative we expect the solution will employ a hybrid approach wherein a cloud hosted Web Portal will provide the interface for users to login and utilize DCC capabilities that span across this web portal, MGHPCC, data Toolsets

and a Snowflake data warehouse supporting AI and data resources. We expect to incorporate open source tooling where possible fostering community engagement and lower costs of implementation and ownership.

Quality and Clarity of Proposed Technical Architecture

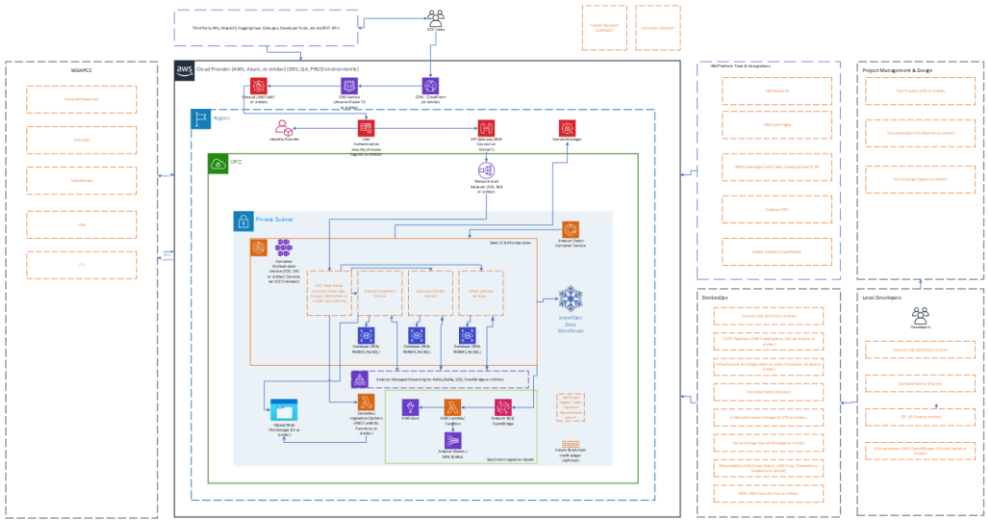


Figure 16: *Sample physical architecture diagram with a potential set of components and open variable components*

As evidenced above, we have thought through the building blocks necessary to deliver a leading class platform that would excel in meeting the data commons collaborative requirements. Taking into consideration the high-level nature of those provided requirements and the resulting variability of implementation options, EY recognizes considerable blind spots that must first be addressed to arrive at and present a high-quality, clear and complete technical architecture which aligns with Massachusetts Technology Collaborative's stated goals.

Failure to adopt this measured approach risks inappropriate architecture and technology choices that either underserve functionality and long term platform viability or incorporate expensive, underused tools that overcomplicate future platform extensibility and maintainability. The result in both scenarios would be extensive re-work and added costs.

In comparison, taking the critical first step with our Discovery phase will enable necessary collaboration between Massachusetts Technology Collaborative stakeholders and EY professionals to define the future state and create a well aligned, high-quality architecture that is not only the correct fit, but more cost effective and efficient.

Resumes & Bios

Qualifications and availability of proposed team

Below we provide indicative profiles for EY personnel who may be assigned to the MassTech AI Hub DCC engagement for each key leadership role. Due to the non-deterministic nature of RFP procurement and contracting timelines we cannot guarantee the availability of the resources below. However, with 400,000 professionals worldwide and over 2,300 professionals living and working in Massachusetts, we are positioned to rapidly respond to MassTech AI Hub engagement needs with resources of similar experience and qualifications if the resources below become unavailable.

Matt Allen

Education

B.S., Computer Science- Saint Michael's College

Certifications

Certified Scrum Master

Project role: Program Lead, Overall

Matt Allen is a Managing Director based in the Boston office, working within the Tech Consulting organization within the Government and Public Sector practice of Ernst & Young LLP.

Matt has more than 25 years of experience providing technology strategy and implementation services to clients across the US, with a concentration in the Northeast. He has provided these services across multiple sectors, with an emphasis on State and Local Government clients.

The initiatives he has helped lead and advice that he provides to clients include application design and development, data strategy, cloud migration, collaboration, and change management, with a focus on Microsoft technologies. Matt has delivered engagements to mid-market and enterprise clients in both state & local government and commercial around the country. Matt has experience leading both onshore and offshore teams.

Relevant experience

- Matt has led program delivery from strategy, to design, to implementation, of complex technology initiatives for several state agencies in Massachusetts, Rhode Island, Connecticut, and Maine over his years of experience. In Massachusetts Matt has successfully delivered transformative solutions for the following state agencies:
  - Executive Office of Housing and Livable Communities
  - MassHousing
  - Trial Court
  - Committee for Public Counsel Services
  - Office of the State Auditor
  - The Center for Health Information and Analysis
  - Executive Office of Labor and Workforce Development
  - State Ethics Commission
  - Massachusetts Gaming Commission
  - Cannabis Control Commission
  - Boston Public Health Commission
- Matt has been the engagement executive supporting a multi-year application modernization program at the Massachusetts Executive Office of Housing and Livable Communities (EOHLC) since 2017. This program allowed the Office to manage a 6x increase in project demand, shifting from paper-based to digital processing of project details and improve external communication with third-party project resources. The improvements made to business processes and technology allowed the Department to continue working throughout the COVID-19 pandemic, something that would not have been possible without this modernization effort. The relationship with EOHLC began in 2017 and continues today.

Morgan Wildman

Education

B.A., University of Manitoba

Certifications

Project Management Institute – Agile Certified Professional (PMI-ACP),
SAFe 6.0 Scrum Master (SSM),
Project Management Professional (PMP),
Prosci–Delivering Project Results

Skills

Project Management:
agile, waterfall, blended.
Integration Management.
Project/Program Governance.
Agile delivery, Scrum, Kanban.
Requirements Gathering.
Scope, Cost, Schedule, Risk, Quality Management.
Excel (expert knowledge).
JIRA, Rally, Azure Devops.

Project role: Project Manager

Morgan Wildman is a Senior in Technology Consulting at EY with 16 years of leadership experience in project and program management spanning all SDLC phases and the product life cycle. Collaboratively addressing organizational needs and objectives with contagious enthusiasm and a reliance on data supporting informed decision making. Skilled in translating objectives into requirements and in addressing evolving project requirements through agile, waterfall, and blended methodologies. Prior to consulting he was a Canadian Forces Army Reserve Logistics Officer and a university student.

Relevant experience

Agile / Hybrid Project Management

- Program PMO, Team Coach / Scrum Master, for a collection of Generative Artificial Intelligence technology accelerator incubation projects for an Internal Advice engagement for EY Global, facilitating lean portfolio management with weighted shortest-job first prioritization, streamlining reporting to Partners and EY Americas, drafting SOWs and pricing plans, supporting ETCs, billing and invoicing, and coaching two of the incubation teams in agile ways of working.
- PMO and Project Manager, of an SAP Commerce Cloud Implementation for a leading MedTech client adopting, adapting and skilling-up agile ways of working by collaborating vertically across ranks and horizontally across teams to promote timely incremental delivery of committed scope at desired quality and cost. Managed program risks and issues, gathering and presenting clean PowerPoint status reports to client directors and executives enabling constructive conversations toward identifying and executing solutions. Facilitated team retrospectives and shepherded continuous improvements to agile ways of working resulting in predictable and sustainable release cycles.
- Project Manager / Scrum Master / Product Manager for a medium insurance client's cloud and analytics development team delivering enhancements across several internal and external vendor teams, and small video game studio client developing licensed and original intellectual property products and features, gathered and translated business objectives, requirements and inputs from key stakeholders at all management levels to identify gaps and inefficiencies then developed and delivered culturally understood and accepted processes, policies, and systems enhancements avoiding or preventing previous performance issues.

Organizational Change Management / Agile Transformation

- PMO Designer for multiple organizations including a medium insurance client, large health care client, large OEM automotive client, and small video game studio client, assisted in development and delivery of an agile transformation roadmap, adoption of agile processes, and documentation of agile terminology. Coached Directors, Program Managers, Product Owners, Product Managers and Team Members in agile methods and mindsets. Supported organizational change management by leveraging Prosci's ADKAR model for change. Coached 60+ teams on how to provide greater business value, drastically reducing meeting waste in duration and frequency.
- Project Lead for an Automotive OEM client's Enterprise Technology Renewal and Cloud Adoption Program on the portfolio team for 2,300+ application teams, supporting Marketing & Sales and Customer Service Division provided expertise in adopting agile practices and strategic advice on cloud adoption, and streamlined communication and reporting across 2,300+ teams.

Waterfall / traditional project and program management

- Prior to EY, Morgan served from 2003-2013 in the Canadian Forces Army Reserves, Land Force Western Area as a Logistics Officer, where he planned and lead increasingly complex projects culminating as the Service Battalion's S3 (Operations & Training Officer).

Matt Galvin

Education

B.S., Computer Science, NY CUNY

Skills

Large-scale system architecture design; Microservices architecutres; Event Driven; Content Management; Scalable solution design; AWS, Azure, Azure DevOps, Docker, Kubernetes, GitHub Actions

Project role: Solution Architect Lead

Matt Galvin is a Senior Manager in the Technology Consulting practice of Ernst & Young LLP. Matt has 25+ years of experience as a dedicated, proven, hands-on, full stack and full lifecycle Digital Transformation Leader, Technical Development Lead, Cloud Solution Architect and Technical Writer providing holistic team leadership and project implementation across a broad range of industry sectors. Matt is driven to hit the launch target while being a team mentor and filler-in of the blanks. His areas of specialty include Solution Architecture, scalable event driven microservices commerce platforms, DevOps, Content Management Systems along with cross-platform Mobile Application design and implementations. Matt's deep background across technologies and industries allows him to bring unique perspectives and thought leadership that comes from hands-on experience and exposure to business needs of all shapes and sizes from planning and inception to implementation, launch, training and long view managed support of product life cycles in the hands of real users that drive business value and revenue.

Relevant experience

- Multiple government, entertainment & education industry clients – Large Scale Content Management Solution Designer: Matt has supported the planning and implementation of multiple CMS platforms with eCommerce, Mobile Applications, server architecture, DevOps and employee training services at scale size clients that are used by constituents, customers & students globally.
- Global oil & gas client - Cloud solution architect: Matt has recently assisted a large global client in the solution architecture definition for a large-scale commerce platform digital technology modernization involving the migration of a legacy platform to a modern cloud event-driven architecture solution.
- Multiple national quick service restaurant clients – Event driven commerce platform architect: Matt has recently assisted in delivering and commercializing customer-facing hybrid mobile stack solutions driven by APIs, event-driven microservices, and integrated observability that are client's primary revenue generation platforms managing millions of transactions a day and billions of dollars in revenue annually. His deep understanding of full-stack platforms development, combined with his leadership skills, has been instrumental in achieving these milestones.
- International cruise line hospitality client – DevOps Platform Engineering Lead: Matt has supported the DevOps architecture and implementation of scale sized platform solutions for a high-profile hospitality industry client revenue generating brand web sites, ecommerce and mobile solutions and cross brand platforms.

Essex De Guzman

Education

B.S., Computer Science, Univ of N. Carolina, Charlotte

Project role: Data Architecture Lead

Exequiel "Essex" De Guzman is a Senior Manager Consultant in the Data & Analytics service part of the National Advisory practice specializing in Emerging Technologies, Data Engineering, Data Architecture, Data Strategy and Data Delivery, Product Conception: Design & Development, Hadoop (Big Data), Data Lake Development, Data Warehousing, Data ETL & Integration, Cloud Architecture and Implementation, and Data Platform modernization.

He has 17+ years of creating and executing an end-to-end data & technology strategy and delivery, from application data creation and data management, to analytical data consumption, and data science model integration into business workflow. A trusted data advisor that guides transformation of data into customer value by applying industry best practices and frameworks to understand what's possible, understanding the business context to determine what's feasible, and implementing enterprise grade solutions to make things real.

Relevant experience

- Engineering Lead (Data, Software and Quality) – Essex owned the data strategies and target architecture for the Sales, Service, Acquisition and Risk Engineering value stream for the middle and large commercial business lines for one of the largest commercial insurance providers. He was responsible for executing an organizational transformation to fine-tune, mature, and grow teams

for the Software, Data and Quality engineering practice while driving down overall cost. He collaborated with executives and delivery leads to modernize the data ecosystem from on premise appliances to hybrid cloud services for data integration, data analytics and data science processes. He provided reverse mentoring for senior executives on emerging technology and modern workplace trends. He created a data hub strategy for application and data integration to accelerate application rationalization and modernization efforts for the organization.

- Data Strategy Lead – Essex was responsible for defining, executing and maintaining one of the largest commercial insurance provider's enterprise data strategy, spanning all lines of business. He established a capability maturity framework to prioritize technology investments in alignment with business needs. He created the data operating model for the business information architects, data architects, platform enablement team, data management services team and delivery team to guide investment, prioritization and acceleration of data ecosystem modernization and simplification initiatives. He defined and implemented the 3PD Hub framework and Architecture, first release was an enterprise firmographic data enrichment API service. He defined and implemented the data management services API orchestration of metadata, data quality, standardization, cleansing, RDM and MDM to accelerate overall delivery and improving data trust. He sponsored and chaired the data architecture review board, sponsored and managed the data technology proof of concepts, and owned the data domain as part of the enterprise technology life cycle management process. He introduced and promoted the adoption of DAMA framework to the enterprise data office. He established the DataOps guild to promote adoption of data DevOps practices. He Established the business information and data architecture community of practice to promote standards and best practices across the organization.
- Big Data Software Engineering Manager – Essex was a manager for 1 of the Big Four consulting firms. He was responsible for rapidly architecting, designing, prototyping, and implementing data systems. He consulted on the Big Data and Data Science needs for a variety of Fortune 1000 corporations and other major organizations. He was responsible for designing, planning and presenting strategic road maps for data initiatives to clients. He worked in cross-disciplinary teams within the consulting firm's industry experts to understand client needs and ingest rich data sources such as social media, news, internal/external documents, emails, financial data, and operational data. He researched, experimented, and utilized leading Big Data methodologies such as Hadoop, Spark, Netezza, and NoSql. He architected, implemented and tested data processing pipelines and data mining / data science algorithms on a variety of hosted settings, such as AWS, Azure, client technology stacks, and the firm's internal clusters. He translated advanced business analytics problems into technical approaches that yielded actionable recommendations in diverse domains such as risk management, product development, marketing research, supply chain, and public policy; communicated results and educated others through insightful visualizations, reports and presentations.
- Technical Architect – Essex was the technical architect for a healthcare informatics organization working on the Enterprise Platform Analytics team, whose goal is to provide data warehousing as service on the cloud by collecting and organizing data from a multiple member hospital's electronic recording system into a data lake. Then, transformed and integrated data into a standard input format (SIF) that feeds into a platform that transforms data into a standard output format (SOF), using a unified proprietary data model and big data technologies such as Scala, Spark, Hive and Oozie. SOF data was then standardized using reference data management (RDM) tool - Symedical, and data entities through a master data management (MDM) process using IBM Initiate. Enriched data was then used for reporting and analytics using BI tools such as Tableau and Cognos. His responsibilities included:
    - Architect, design and develop software and major functionalities of the system.
    - Coordinate with architects and business experts to translate key strategic objectives into actionable and governable roadmaps and designs.
    - Analyze business and technology challenges, assess costs, and suggest solutions.
    - Design and execute proofs- of- concept for critical enterprise capabilities, including building business cases with costs, benefits, and efficiency measures.
    - Architect and build an ETL system to process raw data arriving from the hospitals using Scala and Spark and stored in HDFS, applying complex transformations along the way.
    - Create Oozie workflow schedules and scripts to automate deployments (integration with Jenkins).
    - Introduced process standardization and build tools for version control and automated code builds and deployment.
    - Worked on performance improvements of the entire process to meet rising client base.

- Mentored technical teams.
- Evaluated new solutions and technologies.

Matt Donahue

Education

M.S., Human Factors in Information Design, Bentley Univ

B.S., Electronic Media, Arts, and Communication, Rensselaer Polytechnic Institute

Certifications

Skills

Large-scale system architecture design; Microservices architecutres; Event Driven; Content Management; Scalable solution design; AWS, Azure, Azure DevOps, Docker, Kubernetes, GitHub Actions

Project role: Experience Design Lead

Matt Donahue is a Manager in EY's Government & Public Sector practice, based in the Boston office, where he leads user experience strategy and delivery across complex public initiatives. With a passion for crafting intuitive digital experiences and fostering strong client relationships, Matt has recently spearheaded the development of a comprehensive roadmap for the Commonwealth of Massachusetts' Executive Office of Labor and Workforce Development, an effort that engaged over 260 stakeholders and continues into implementation under his leadership. He also served as the lead UX designer and front-end development coordinator for the New Mexico Legislature's custom bill drafting system, translating stakeholder insights into actionable design and technical execution.

Relevant experience

- Roadmap Project Manager: Matt was the project manager for the development of a comprehensive roadmap for the Commonwealth of Massachusetts's Executive Office of Labor and Workforce Development. This roadmap was built over the course of 16 weeks with research participation by over 260 stakeholders. This cross-functional team pulled from many of EY's varied consulting teams including business transformation, people consulting, and technology consulting. Matt also took on the role of coordinating between two additional workstreams at that same client. A subset of the roadmap team was requested by the client to extend their work in order to begin implementing the roadmap.
- Matt was the lead UX designer and oversaw front-end development for the New Mexico Legislature's custom-built bill drafting and workflow system. He conceptualized designs based on stakeholder interviews and a vast history in the legislative world. He wrote requirements, validated those requirements, and oversaw the front-end components (HTML/CSS) of development so that the implementation met those requirements in a way that was true to the design.
- Matt joined EY in 2023. Prior to joining the firm, he worked in Human Centered Design consulting in the Northeast where he gained over 13 years of experience in user experience design and front-end development for government and commercial clients. In his tenure at Tallan, he was the lead designer on legislative websites that combined for three (2013, 2018, and 2021) NALIT Online Democracy Awards, recognizing the best legislative website in the country.

References & Quals

Strength and Relevance of References and Prior Work

Below we provide 3 references substantially similar to the engagement model and outcomes we have proposed as our Discovery Phase 0 in the Approach section. To further demonstrate the breadth and depth and relevance of our prior work, we have included additional customer case studies in Appendix C – Related Experience.

We request that before any reference is contacted directly that the MassTech AI Hub coordinate outreach through EY

References

Reference #1

Contact Name: Michael (Chase) Martin

Organization: Federal Bureau of Investigation

Email Address: mcmartin@fbi.gov

Phone Number: 504-220-9777

Project Description: Over the past 3 years, EY has assisted the internal audit section of the Inspections Division (INSD), formerly known as the Office of Internal Auditing (OIA), by developing and managing a cloud analytics solution with an FBI system security authority to operate (ATO). Through this solution, EY provides analysis and reporting on the Bureau's national security activities, namely use of the Foreign Intelligence Surveillance Act (FISA) and querying of FISA-collected intelligence. The EY team enables the FBI to provide trusted and easily compiled analyses to both internal stakeholders (across all FBI Field Offices (FOs), headquarters divisions, and 7th floor leadership) as well as external (DOJ, Congress, and the Office of the Director of National Intelligence (ODNI)). EY's contributions were instrumental in the reauthorization of FISA Section 702, which included direct briefings to FBI leadership and compilation of materials used in congressional briefings.

EY has developed Python models for data classification, user behavior analysis, and compliance rick scoring to support OIA/INSD in triaging national security activity review. These models process millions of records, classify them into actionable categories, and calculate risk scores for FISA queries by combining results from multiple analytic tests. The scoring logic is continuously updated in collaboration with OIA and NSCLB to align with evolving policies. EY also assists OIA/INSD in leveraging AI tools within Microsoft Azure Secret Cloud, preparing business case briefings for the FBI OCIO, and advancing AI applications through the EY developed Query Assistant training platform. In addition, EY manages a robust analytics ecosystem through Power BI dashboards ("Navigators") deployed across all FBI Field Offices to provide FISA querying insights to legal directors and staff. EY engineers maintain secure cloud infrastructure, orchestrate data pipelines via Azure Data Factory, and implement SQL/Python logic for integration and quality checks. EY also has developed React.js-based tools like Query Assistant to help FBI personnel craft compliant queries, while ensuring governance, documentation, and system updates through FBI technology control and security processes.

Reference #2

Contact Name: Cynthia Raulerson

Organization: Department of Defense

Email Address: cynthia.a.raulerson.civ@mail.mil

Phone Number: 703-692-9131

Project Description: The Office of the Undersecretary of Defense (Comptroller) enlisted EY to enhance cost and performance management within the Department of Defense by developing data analysis and visualization tools. EY initially focused on categorizing Operations and Maintenance expenses into seven business lines, later expanding its efforts to the Revolving Funds and RDT&E portfolios. In the Revolving Funds area, which handles over $120 billion annually, EY introduced automated dashboards that track cash balances, net operating results, and fund statuses. These tools enabled senior leaders to proactively manage resources and improved mid-year budget reviews through automation, creating a feedback loop between budget planning and execution that could be replicated across other funding areas.

In the RDT&E domain, EY applied advanced analytics, including natural language processing, to categorize over $100 billion in annual spending into critical technology areas. This allowed the DoD to better link procurement and budget data to actual obligations and expenditures, supporting more informed decision-making. EY also contributed to initiatives like the Technology Tracking Action Group and the Rapid Defense Experimentation Reserve. Across all efforts, EY's strategic planning and analytics, particularly through the Advana Enterprise Analytics platform, helped the DoD improve visibility into financial data, refine strategies, and enhance organizational efficiency while mitigating risks.

Reference #3

Out of respect for our client's privacy we are unable to provide direct contact information for our third reference, however we will facilitate a reference meeting upon request at a date and time that is mutually available. Please coordinate contact with this reference through Christopher Davidson, contact information below.

Contact Name: Christopher Davidson **Organization:** name withheld for confidentiality

Email Address: christopher.davidson@ey.com **Phone Number:** 215-448-5497

Project Description: Our client is well-known for their personalized, efficient, and reliable in store experience. With changing customer demands, they knew they needed to scale their footprint onto digital platforms, becoming an omnichannel company while also creating the ability to seamlessly launch new lines of business offerings. Our client wanted to build a digital strategy and roadmap to guide their digital transformation priorities. They needed help organizing ideas coming from multiple sources and at varying levels of maturity and criteria of ”success.” Our EY team helped the client lay out a market-competitive tactical roadmap for the entire digital program with milestone dates and corresponding realistic levels of effort to complete each milestone. We also helped them think through the future of IT in 3-5 years and how IT could help enable their business’ objectives.

EY brought proven industry leading expertise around system architecture, business strategy, commerce, engineering, product management, and SAFe execution, which was embedded across 10+ focused product teams. The team built milestone roadmaps for the current digital initiatives in-flight and for the remainder of the client's transformation scope and beyond, creating transparency among leadership and aligning agile teams on the goals and vision of the digital program. We designed a 3-5 year technology roadmap for the digital program to achieve objectives of the business by expanding the platform they have today and introducing disruptive bleeding edge tech. At the end of the project, the team established a quantifiable release management process using metrics and level of effort to manage scope / budget focusing agile teams around product goals and release timelines. EY & our client together implemented and launched a collection of new in-store & online solutions as well as launching a new line of business offering on a digital platform to 945+ locations, decomposed and reimagined CI/CD pipelines and provided a new IaC provisioning mechanism, implemented a consistent, automated end-to-end testing and UI test automation enables early detection of issues, reducing bugs in production and coached client leadership and agile teams in leading SAFe teams and adopting agile tools, technology, and ways of working enabling our client to be able to maintain and enhance their platform on their own terms and grow into the future.

Cost Response (Attachment C)

To provide Massachusetts Technology Park with a predictable cost structure EY is proposing a firm-fixed-price for the Discovery phase.

As introduced in the Approach section and further discussed in the section Feasibility of Timeline to Complete Program Scope, our proposal is based on a 16-week timeline. This assumes the articulated time commitments from key workstream stakeholders as outlined in that section.

Please find EY's Solution Pricing Template for the 16-week cost model attached to this bid under the filename *Services Budget Template.xlsx*. Our proposed cost is subject to the assumptions outlined in Appendix A – Proposal Assumptions.

We are pleased to proactively offer for your review an alternative, compressed 12-week timeline as outlined in Appendix B – Compressed Timeline. Mentioned in the section Feasibility of Timeline to Complete Program Scope EY is prepared to move through the Discovery phase more quickly should Massachusetts Technology Collaborative feel confident they can dedicate additional time each week to move through the outlined activities at an accelerated pace. Our activities, deliverables, outcomes, and success would remain the same under the compressed timeline. Should you wish to entertain this alternative EY would be happy to provide a revised cost response in alignment with a 12-week schedule.

Budget realism, alignment with scope, and cost-effectiveness

EY's multidisciplinary team, in collaboration with our alliance ecosystem, evaluated the scope outlined in the RFP, Data Commons Collaborative Requirements, and provided Q&A. It became readily apparent that the appropriate place to start was with a Phase 0 Discovery engagement. The optionality inherent in the broadly defined requirements leaves much room for interpretation on how wide and deep we could go to deliver a solution for Massachusetts Technology Collaborative.

As we ruminated on potential solution options such as open source tools and associated customization needs, security protocols, commercial products and their customization needs, yet to be articulated business process complexity and workflow steps, infrastructure redundancy, failover and disaster recovery, web portal options, meta-data tooling options and capabilities, DevOps processes and tooling, details of the credit system, and overall user experience, all along with associated pros and cons on each decision point, we became increasingly certain that proposing a Discovery phase in alignment with the Initiation & Design Phase outlined in the DCC Requirements.PDF document (captioned below) was advisable.

Phase	Milestone	Description
1. Initiation & Design	Project Kickoff	Core team resourced, aligned, scope and goals confirmed, project charter approved
	Detailed Architecture & Governance Defined	Detailed technical design and initial governance policies finalized
	Dataset Curation Criteria Approved	Criteria for dataset inclusion reviewed and approved by stakeholders
	Initial Dataset List Confirmed	Validated preliminary list of hosted and external datasets

Figure 17: *Excerpt from DCC Requirements Timeline*

Based on the potential variability in detailed requirements gathering, our experience informs us that the investment required for the Data Commons Collaborative solution could fall anywhere in the broad range of $10-50m. Rather than furnish a proposal based on high-level business requirements and assumptions, we concluded we should first work to define requirements, refine scope, outline success criteria, and collaborate to arrive at a solution that will effectively meet the needs of Massachusetts Technology Collaborative in a cost-effective manner. Adopting this path will empower the Collaborative to understand its options and align priorities with budget realities.

EY's reputation and brand are predicated on serving our customers as trusted advisors. With a track record of success in the Commonwealth, across the US, and globally, we encourage Massachusetts Technology Collaborative to consider our proposal as a measured, realistic, cost-conscious approach that stands above that of competing vendors.

In Conclusion

In embarking on a project of this magnitude, it is crucial to undertake a thorough and focused Discovery, Design, and Planning engagement. This initial phase is essential for refining the architecture and functional capabilities that will not only meet the stated goals but also establish

a modular foundation for future growth. By engaging in this detailed planning process, we can design a solution that is secure, scalable, and adaptable, with predictable costs and an achievable timeline. This approach allows us to clearly define and measure goals, specify detailed functional requirements, select compatible and capable technology components, and develop an actionable, phased implementation plan. Ultimately, this discovery engagement will lay the groundwork for a successful implementation, providing a comprehensive understanding of cost and effort, and equipping the MassTech AI Hub with an immediately actionable roadmap for the future.

At the conclusion of this engagement, we will hand over the following assets which will allow for a seamless transition from planning into implementation:

- Functional Requirements – a backlog of specific and detailed functional requirements that are "shovel-ready" including core system capabilities, functional actions, controls, governance, fairness, workflows, interaction elements and more
- UI/UX Design – visual assets that represent the future state of the web-portal, along with the logic behind the crediting system
- End to End Physical Solution Architecture – visual and narrative description of the future state end to end physical architecture, accommodating for a hybrid deployment
- End to End Logical Technology Architecture – solution specific diagrams that reflect the technical components that comprise the overall solution
- End to End Data Architecture – a specific and detailed diagram representing data flow, data residency, and workload integrations
- Infrastructure, DevOps and Security Architecture – a set of solution documents outlining the Infrastructure-as-Code, CI/CD deployment, and overall Security Architecture for the platform, with component selection, interaction details, and maintenance information
- Implementation Workstream Roadmap – an achievable, actionable implementation roadmap covering all workstreams through a coordinated team
- Program Plan & Governance Structure – a critical aspect of delivery, this will be the structure that enables proper project management, stakeholder engagement, and progress visibility

With the above information in hand the MassTech AI Hub will be equipped to immediately implement the solution from the foundation layer on up to the web portal over a phased, achievable timeline.

As individuals who serve, and in many cases live in, the Commonwealth, we are excited about the DCC platform that the MassTech AI Hub will create, and we want to be a part of that story. Massachusetts has always been a beacon for revolution, and that story is carried on through technological advancements that start right here. This platform will democratize access to research data that will spur a new generation of innovation. We are here to provide the technical depth of experience, and strength in complex project and program management, to bring this platform to life.

We appreciate your considering our proposal and welcome any opportunity to further discuss our approach.

W-9

Form **W-9**
(Revised April 2022)
Massachusetts
Substitute Form W-9

Request for Taxpayer Identification Number and Certification

▸ **Online instructions at:** macomptroller.org/wp-content/uploads/instructions_w-9.pdf

Give this Form to the requestor or the department you are doing business with.

Print or type. See Specific Instructions on page 3.

1 Business name/Taxpayer (as shown on your income tax return). Name is required on this line; do not leave this line blank.
Ernst & Young U.S. LLP

2 Business name/disregarded entity name/dba, if different from above.

3 Check appropriate box for federal tax classification of the person whose name is entered on line 1. Check only **one** of the following seven boxes.

☐ Individual/sole proprietor or single-member LLC ☐ C Corporation ☐ S Corporation ☑ Partnership ☐ Trust/estate

☐ Limited liability company. Enter the tax classification (C=C corporation, S=S corporation, P=Partnership) ▸ _____

Note: Check the appropriate box in the line above for the tax classification of the single-member owner. Do **not** check LLC if the LLC is classified as a single-member LLC that is disregarded from the owner unless the owner of the LLC is another LLC that is not disregarded from the owner for U.S. federal tax purposes. Otherwise, a single-member LLC that is disregarded from the owner should check the appropriate box for the tax classification of its owner.

☐ Other (see instructions) ▸

4 Exemptions (codes apply only to certain entities, not individuals; see instructions on Page 4):

Exempt payee code (if any): _____

Exemption from FATCA reporting code (if any): _____

(Applies to accounts maintained outside the U.S.)

5 Legal Address (number, street, and apt. or suite no.) See instructions.
One Manhattan West

6 City, state, and ZIP code
New York, NY 10001

Requester's name and address (optional)

7 Remittance Address (if different from Legal Address)
200 Plaza Drive, Suite 2222, Secaucus NJ 07094

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on line 1 to avoid backup withholding. For individuals, this is generally your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the instructions for Part I, on Page 5. For other entities, it is your employer identification number (EIN). If you do not have a number, see *How to get a TIN*, on Page 5.

Note: If the account is in more than one name, see the instructions for line 1. Also see *What Name and Number to Give the Requester* for guidelines on whose number to enter.

Social security number
___ - ___ - ____

or

Employer identification number
34-6565596

DUNS Number
Please confirm with the state agency if this is required for vendors receiving federal funds.

Unique Entity Identifier (SAM)
As of April 4, 2022, all vendors that receive federal grant funds must submit their Unique Entity Identifier registered in the System of Awards Management (SAM).

Part II Certification

Under penalties of perjury, I certify that:
1. The number shown on this form is my correct taxpayer identification number (or I am waiting for a number to be issued to me); and
2. I am not subject to backup withholding because: (a) I am exempt from backup withholding, or (b) I have not been notified by the Internal Revenue Service (IRS) that I am subject to backup withholding as a result of a failure to report all interest or dividends, or (c) the IRS has notified me that I am no longer subject to backup withholding; and
3. I am a U.S. citizen or other U.S. person (defined below); and
4. The FATCA code(s) entered on this form (if any) indicating that I am exempt from FATCA reporting is correct.

Certification instructions. You check the following box if you have been notified by the IRS that you are currently subject to backup withholding because you have failed to report all interest and dividends on your tax return. For real estate transactions, item 2 does not apply. For mortgage interest paid, acquisition or abandonment of secured property, cancellation of debt, contributions to an individual retirement arrangement (IRA), and generally, payments other than interest and dividends, you are not required to sign the certification, but you must provide your correct TIN. See the instructions for Part II, on Page 5.

☐ Item 2 does not apply.

5. I am an active Commonwealth of Massachusetts state employee: (check one) ☐ Yes ☑ No

If yes, I certify compliance with the Massachusetts State Ethics Commission requirements at https://www.mass.gov/ethics.

Sign Here | **Signature of U.S. person** ▸ DocuSigned by: 2A365CBE7A694E5... | **Date** ▸ 06/14/2024

Massachusetts Substitute Form **W-9** (Rev 4-2022)

Signature Acceptance Form (Attachment B)

Attachment B

Massachusetts Technology Collaborative

Authorized Respondent's Signature and Acceptance Form

The undersigned is a duly authorized representative of the Respondent listed below. The Respondent has read and understands the RFP requirements. The Respondent acknowledges that all of the terms and conditions of the RFP are mandatory, and that Respondent's response is compliant with such requirements[1].

The Respondent understands that, if selected by the Mass Tech Collaborative, the Respondent and the Mass Tech Collaborative will execute an Agreement specifying the mutual requirements of participation.

Respondent agrees that the entire bid response will remain valid for sixty (60) days from receipt by the Mass Tech Collaborative.

I certify that Respondent is in compliance with all corporate filing requirements and State tax laws.

I certify that the statements made in this response to the RFP, including all attachments and exhibits, are true and correct to the best of my knowledge.

I acknowledge and understand the procedures for handling confidential materials submitted to the Mass Tech Collaborative as set forth in Section 3.1(d); I agree to be bound by those procedures and agree that the Mass Tech Collaborative shall not be liable under any circumstances for the disclosure of any materials submitted to the Mass Tech Collaborative pursuant to this RFP or upon the Respondent's selection.

Respondent: Josh Liberman

By:

Name: Josh Liberman

Title: Authorized Signatory

Date: 10-2-2025

[1] Subject to discussion and mutual agreement of final terms.

Appendix A – Proposal Assumptions

As part of our response to the RFQ, we have reviewed the terms and conditions set forth therein. While we are willing to enter into an engagement agreement based upon the proposed terms, our proposal to provide services is contingent upon the execution of a mutually satisfactory engagement agreement. Our proposal is conditioned on finalizing the terms included in a draft Statement of Work and upon award we will work with MassTech to expeditiously finalize the Statement of Work.

Please see additional assumptions below:

General Project Assumptions:

1. Engaging EY in the Discovery Phase will not be precluded us from engaging for the implementation phases of the program.
2. Key stakeholders will be available to meet with EY team members on a regular basis for requirements gathering meetings, workshops, reviews, checkpoints, planning, and status. We anticipate that at least one key stakeholder from each of the 7 workstream pillars described in the proposal will need to allocate 15 to 20 hours per week, on average, throughout the 16-week engagement. If these key stakeholders do not have availability, the project schedule and cost will be impacted and require a change order.
3. EY expects to meet on-site regularly with project stakeholders throughout the engagement. It is expected that key stakeholders will be available to meet with EY resources in person. Lack of in person engagement will impact efficiency and create schedule delays which will necessitate a change order.
4. Feedback on deliverables will be supplied within 5 business days. Should comments not be provided within this time frame, the deliverable will be deemed accepted by default.
5. When EY resources are on-site, the Massachusetts Technology Collaborative will provide a satisfactory physical workspace for the EY consultants.
6. The proposed costs contained in our proposal are for the Discovery Phase only. It does not include implementation, testing, deployment, or training on any solution.
7. Engagement deliverables are limited to information gathered and incorporated during the fixed 16-week schedule. Additional needs beyond 16 weeks will require a change order.
8. Delays in data access or approvals may impact delivery timelines and cost.
9. Deliverables will not include EY branding.

Scope Assumptions:

1. The proposed costs for the Discovery phase approach as outlined herein are based on the requirements areas outlined in the RFP and DCC Requirements.PDF. Any functionality not mentioned in the RFP or DCC Requirements.PDF is out of scope for Discovery by default. If additional scope items are desired for inclusion, it will require a change order.
2. All existing documentation, proof of concept source code, and environment details will be provided to the EY team upon project kickoff.
3. Read-only access to MGHPCC will be granted to facilitate execution of Discovery phase activities.
4. Documentation pertaining to MGHPCC including design documentation, logs, metrics, incident records, deployment history, etc. will be made available to EY personnel.
5. Approvals and access requests for EY team members will be completed on schedule.
6. Security requirements identified will be limited to the systems and facilities specified in the project scope and will not extend to other areas unless explicitly agreed upon.
7. MassTech to provide defined incident response protocols that will be integrated into the security monitoring and alerting requirements defined during the project.
8. MassTech to provide clear definitions and classifications for data sensitivity, which will guide the development of data protection and classification requirements.
9. Regulatory, security, and data residency requirements (e.g., HIPAA, GDPR, state mandates) will be provided by MassTech and will be used to guide design decisions.

10. A single primary cloud provider will be identified and approved as an outcome of Phase 0.
11. Existing on-prem infrastructure details (e.g., compute, storage, networking) will be available and stable for integration with the cloud.
12. Hybrid Connectivity – Reliable and secure connectivity (e.g., VPN, Direct Connect, ExpressRoute) will be provisioned by MassTech within project timelines and will meet baseline performance needs.
13. Business, security, and infrastructure stakeholders will actively participate in workshops, requirement reviews, and approvals within sprint timelines.
14. Workshops – Workshops will be 4-5 deeper dive collaborative planning sessions enabling detailed understanding of requirements per workstream and dependencies across workstreams. These may be in-person, hybrid or virtual as needed.
15. Review sessions – As each workstream conducts collaboration sessions around requirements and decisions being made, each workstream will have 2-3 review sessions to validate discoveries and reinforce stakeholder alignment to inform selections, requirements and roadmap priorities. These may be in-person, hybrid or virtual as needed.
16. Sign-off Approval sessions – Each work stream will have 1 final sign off session with its workstream leadership stakeholders with 1 additional finial group sign-off workshop which may take the form of a “mini Pre-PI Planning” session enabling finial implementation readiness leading to next steps for the build phase and developer onboarding. These may be in-person, hybrid or virtual as needed.

Appendix B – Compressed Timeline

12-Week Timeline

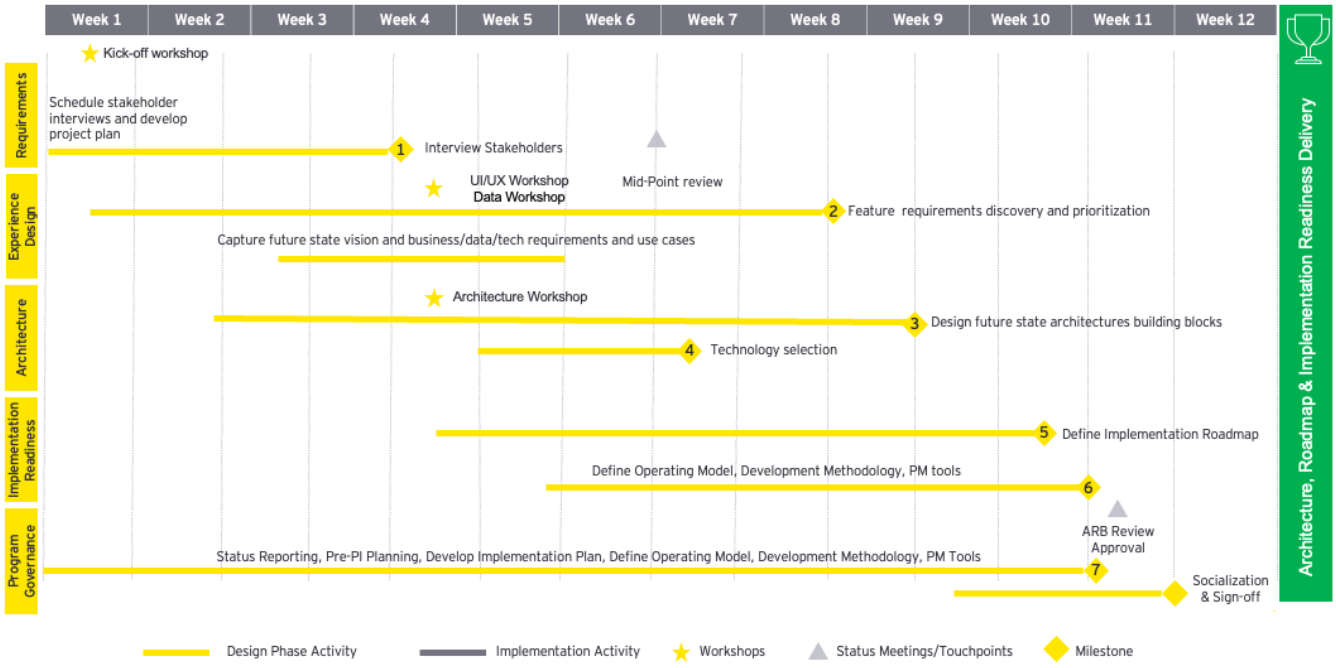


Figure 18: *Proposed timeline of activities and order of operations*

12-Week Activity Schedule

Program & Project Management (PMO)

	Week	Key Activities
Kickoff	Week 1	• Establish kickoff, workshop and collaboration meetings cadence • Establish regular leadership reporting check ins • Establish weekly status reports • Engage stakeholders and schedule workshops • Review activity plan, timeline, and expectations with each team
Sprint 1	Week 2	• Facilitate functional and non-functional requirements capture in PM tools to enable later backlog generation • Facilitate cross team collaboration sessions and attendance
	Week 3	
Sprint 2	Week 4	• Define Ways of Working • Define RACI • Define program operating model • PM tools technology selection and facilitate tools availability
	Week 5	

Sprint 3	Week 6 Week 7	• Facilitate early Agile training and program setup operations PM tools technology selection and facilitate tools availability • Oversee and validate backlog generation
Sprint 4	Week 8 Week 9	• Define training requirements • Facilitate backlog prioritization sessions
Sprint 5	Week 10 Week 11	• Facilitate Pre-PI planning workshops • Lead the prioritized roadmap creation and socialization
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

Experience Design

	Week	Key Activities
Kickoff	Week 1	• *Kickoff, onboarding & stakeholder identification* • *Stakeholder collaboration sessions*
Sprint 1	Week 2 Week 3	• *Establish catalog of UI screens for the solution* • *Establish user persona mapping to understand requirements for:* ○ *Homepage, Dataset Directory, Dataset Detail Pages, Public Datasets, Synthetic Data Tools, Massachusetts Public Datasets*
Sprint 2	Week 4 Week 5	• *Establish user persona mapping to understand requirements for:* ○ *Community Dataset Submission, Admin Dashboard, Additional Requirements, Data Hosting & Access, User Experience Enhancements, Credit Solution, Using MGHPCC compute resources* • *Design workshop* • *Define user journey maps and workflows*
Sprint 3	Week 6 Week 7	• *Being wireframing capabilities capturing initial low fidelity mock-ups for review* • *UI/UX implementation approach and technology selection*
Sprint 4	Week 8 Week 9	• *Define initial backlog of tasks to define high fidelity design and UI components* • *Prioritized backlog by phase*
Sprint 5	Week 10 Week 11	• *Review sessions and design workshop*
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

Security

	Week	Key Activities

Kickoff	Week 1	• Stakeholder interview to understand all system and facility security requirements • Evaluate existing security capabilities & tooling to identify reuse option
Sprint 1	Week 2 Week 3	• Identify data, cloud, data center and code security needs both functional and non-functional • Capture requirements documentation for
Sprint 2	Week 4 Week 5	• Define security monitoring and alerting requirements • Define security checklist for developers • Define encryption and role based access control (RBAC) requirements • Collaborate with DevOps on CI/CD security integrations • Define data protection and classification requirements.
Sprint 3	Week 6 Week 7	• Security tooling technology selection • Define initial implementation backlog • Document procedure to adhere to industry security best practices
Sprint 4	Week 8 Week 9	• Prioritized backlog by phase
Sprint 5	Week 10 Week 11	• Review sessions
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

Cloud & Infrastructure

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding, stakeholder identification • Stakeholder collaboration sessions • Socialize & gain alignment on scope and objectives
Sprint 1	Week 2 Week 3	• Validate functional requirements • Non-functional requirements planning (performance, HA/DR, compliance) • Identify cloud provider & on-prem infrastructure requirements
Sprint 2	Week 4 Week 5	• Review implementation methodology (provisioning, deployments, hybrid practices) • Draft high-level architecture diagrams (cloud landing zone, hybrid connectivity) • Technology/tool validation for services and interoperability
Sprint 3	Week 6 Week 7	• Calculate running hosting & usage costs (based on tools and expected load) • Define initial implementation backlog
Sprint 4	Week 8 Week 9	• Prioritize backlog items by phase (Seed → Grow → Bloom)

Sprint 5	Week 10 Week 11	• Review sessions with stakeholders • Establish governance framework (decision-making, reporting cadence) • Define compliance/security baselines (logging, monitoring, encryption, retention policies) • Draft risk register and mitigation strategies
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Define implementation readiness next steps • Develop training/knowledge transfer plan • Prepare runbooks for Day-1/Day-2 operations • Change management & communication planning

Web Portal & Credit System

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions
Sprint 1	Week 2 Week 3	• UX and user workflow • Detailed functional requirements for: ○ *Homepage, Dataset Directory, Dataset Detail Pages, Public Datasets, Synthetic Data Tools, Massachusetts Public Datasets, Using MGHPCC compute resources* • Non-functional requirements planning
Sprint 2	Week 4 Week 5	• Detail feature capabilities to understand full scope needs • Detailed functional requirements for: ○ *Community Dataset Submission, Admin Dashboard, Additional Requirements, Data Hosting & Access, User Experience Enhancements, Credit Solution, Using MGHPCC compute resources* • Evaluate options to align implementation costs with value drivers • Begin solution architecture definition capturing all system and tools that will encompass the system
Sprint 3	Week 6 Week 7	• Experience design collaboration to size UX implementation & requirements • Portal architecture definition & review continued
Sprint 4	Week 8 Week 9	• Define QA testing practices • Define initial backlog • Technology selection enabling efficient development with a lean towards open source technologies as accelerators and cost savers
Sprint 5	Week 10 Week 11	• Prioritized backlog by phase • Collaborate with DevOps to define local developer SDLC to support efficient onboarding and development • Review sessions

Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

Data & Data Services

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions
Sprint 1	Week 2 Week 3	• Interviews for detailed discovery of initial data capabilities ○ Data hosting and Access, Data Curation Plan, Metadata management, Dataset Discovery and Search, Access Control & Governance

Sprint 2	Week 4
	Week 5
		• Interviews for detailed discovery of initial data capabilities
		  ○ Dataset Publishing & Versioning, Collaboration & Community Features, Data Previews & Exploration, Synthetic Data Generation, AI Models, Commons Credits
		• Detailed data requirements
		  ○ Data Assets priorities
		  ○ Data Services priorities
		  ○ Non-functional data requirements planning
		• Governance Requirements
		• AI Fairness Requirements

Sprint 3	Week 6
	Week 7
		• Design Data Capabilities and Services for enablement and integration
		  ○ Define Op Model/Governance for data assets
		  ○ Data Publishing
		  ○ Data Maintenance
		  ○ Data Subscriptions
		  ○ Etc.
		• Define initial backlog of tasks to define high fidelity data design and data services components

Sprint 4	Week 8 Week 9	• Data implementation approach and technology selection • Prioritized backlog by phase
Sprint 5	Week 10 Week 11	• Review sessions and design workshop
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

DevOps

	Week	Key Activities
Kickoff	Week 1	• Kickoff, onboarding & stakeholder identification • Stakeholder collaboration sessions

Sprint 1	Week 2	• Evaluate tool preferences • Identify data, cloud, data center and code deployment needs both functional and non-functional • Capture requirements documentation
	Week 3	
Sprint 2	Week 4	• Identify data, cloud, data center and code observability needs both functional and non-functional • Define CI/CD pipeline requirements for each deployment target and code feedback loops
	Week 5	
Sprint 3	Week 6	• Technology selection of specific tools, services and facility interoperability • Calculate running hosting & usage costs based on tools and traffic load expectations • Define initial implementation backlog
	Week 7	
Sprint 4	Week 8	• Define local developer SDLC workflow • Prioritized backlog by phase
	Week 9	
Sprint 5	Week 10	• Define local developer SDLC workflow • Review sessions
	Week 11	
Discovery Closeout	Week 12	• Socialize and gain solution approvals • Implementation readiness next steps

Appendix C - Related Experience

The following project-specific activities serve to demonstrate EY's breadth of capabilities with legacy data analysis and enterprise data modernization programs:

Commonwealth of Massachusetts

Executive Office of Labor and Workforce Development

Challenge

The client has engaged EY to assess their current workforce development programs and to develop a strategy and roadmap for designing a future state. This future state aims to meet the needs of the diverse stakeholder group served by EOLWD, enhance access to and utilization of data, and incorporate next-generation technology to modernize the user experience. The scope of this assessment includes evaluating their Jobseeker/Employer portal, Case Management System (Staff Portal), Training Provider Portal, and workforce data flow and management.

Our Solution

EY collaborated with the Massachusetts Executive Office of Labor and Workforce Development (MA EOLWD) to successfully develop a strategy and roadmap for modernization. This initiative encompassed business process optimization, the integration of advanced technology, and the enhancement of user experiences for job seekers, employers, and staff across various workforce systems. The project was executed in two phases: the Current State Assessment and the Future State Roadmap. During the Current State Assessment phase, EY identified key in-scope programs, critical functions, and processes essential for modernization. The team reviewed existing documentation and conducted in-depth sessions to gain insights into significant pain points, risks, and challenges. Following the completion of this assessment, EY formulated the Future State Roadmap, which addressed immediate challenges, established infrastructure goals, provided technology recommendations, and outlined the necessary steps to achieve the envisioned future state

Impact

During the Current State Assessment phase, EY conducted a thorough evaluation of key workforce programs within the Executive Office of Labor and Workforce Development (EOLWD), including Job Quest, StaffView, Data Layer, and Integration Layer. The team reviewed existing documentation available on MassGov SharePoint, as well as training materials from JobQuest, TrainingPro, and MOSES. EY facilitated whiteboarding and deep dive sessions with the Chief Technology Officer, program leads, and other stakeholders to gain insights into current capabilities, objectives, technology stack, and the major risks and challenges associated with the existing programs. Utilizing this information, EY identified specific program areas and opportunities for improvement, assessed current manual processes, and analyzed various architectural components and their interactions. The team also explored tools, technology, and data management practices within the workforce systems. Furthermore, EY evaluated the current state of workforce data flow and management across the in-scope programs, drafting findings that highlighted major challenges and areas for improvement related to data management. The findings were validated with key stakeholders, and a gap analysis along with a risk assessment was presented to the program leadership.

In the Future State Roadmap phase, EY developed a proposed vision for Future State Architecture, outlining goals for both the Near Term (less than 5 years) and Long Term (more than 5 years). This included a prioritized roadmap, a procurement strategy, and an execution outline for the roadmap. EY's recommendations for the Future State focused on establishing a modern, unified infrastructure aimed at simplifying access to employment services for job seekers, employers, and training providers, while enhancing operational efficiency, enabling data-informed decision-making, and ensuring long-term adaptability. The procurement strategy included a build versus buy analysis along with corresponding recommendations. The Future Roadmap execution provided a structured framework for implementing a SaaS-First vision. Additionally, EY formulated a Future State Data Strategy based on foundational pillars, including data collection and ingestion, storage and management, processing and analytics, utilization and sharing, and data governance and culture. This strategy is designed to empower EOLWD with data-driven decision-making, intelligent services, and seamless collaboration across agencies.

Commonwealth of Massachusetts

Massachusetts State Health Information Agency

Challenge

This State Health Information Agency was grappling with a series of challenges tied to its reliance on Exadata. As the volume and complexity of healthcare data surged, the limitations of the legacy system became increasingly apparent. Key pain points included:

- Scalability Issues: Exadata was struggling to keep pace with growing data loads, resulting in performance slowdowns and delayed analytics.
- Outdated Data Pipelines: The existing data pipelines were cumbersome and inefficient, making it difficult to integrate new data sources and adapt to changing business needs.
- Collaboration Barriers: There was a clear need for a platform that could enhance data sharing and collaboration among stakeholders, including healthcare providers, researchers, and policymakers.
- Modernization Imperative: The organization recognized the urgent need to modernize its data stack to leverage advanced analytics, machine learning, and real-time data processing capabilities.

Our Solution

To tackle these challenges head-on, EY proposed a robust solution centered around migrating the organization's data infrastructure from Exadata to Snowflake. The solution encompassed several key elements:

- Cloud-Native Architecture: Snowflake's cloud-based design offered the scalability and flexibility required to efficiently manage large volumes of data.
- Revamped Data Pipelines: EY collaborated closely with the organization to redesign and implement modern data pipelines that seamlessly integrated diverse data sources, ensuring timely and accurate data availability.
- Automated Data Collaboration: The new platform included features that automated data sharing and collaboration, empowering stakeholders to access and analyze data in real-time without manual intervention.
- Advanced Analytics Integration: By harnessing Snowflake's capabilities, the organization could leverage advanced analytics tools and machine learning algorithms to extract valuable insights from their data, driving informed decision-making.

Impact

The transition to Snowflake brought about significant improvements:

- Performance Boost: The new data infrastructure delivered a marked increase in performance, enabling faster data processing and analytics, which translated to improved operational efficiency.
- Scalability and Flexibility: The organization can now scale its data operations effortlessly, accommodating future growth and evolving data needs without the constraints of the previous system.
- Enhanced Collaboration: The automated data collaboration features cultivated a culture of data-driven decision-making, allowing stakeholders to work together more effectively and share insights in real-time.
- Strategic Insights: With advanced analytics capabilities, the organization gained deeper insights into healthcare trends and outcomes, empowering them to make informed decisions that positively impact patient care and operational strategies.

State of Colorado

Medicaid Enterprise System (MES) Integration Platform Solution

Challenge

The Centers for Medicare and Medicaid Solutions (CMS) has required states to migrate their core technology platforms from highly integrated legacy Medicaid Management Information Systems (MMIS) to modularized Medicaid Enterprise Systems (MES) that can be more

easily and cost-effectively modified, upgraded and maintained. Colorado sought to initiate its Medicaid systems modernization efforts by engaging an Enterprise Solutions Integrator (ESI) to lead development of a MES strategy roadmap, program governance, and centralized data hub and operational data store (ODS) for inbound and outbound data transfers.

Our Solution

EY was selected as ESI to plan, design, develop and implement the MES Integration Platform Solution for 15+ planned MES modules with the specific goal of replacing over 200 legacy system point-to-point interfaces with modern API capabilities.

The implementation approach includes:

- Developing MES Roadmap with module-specific integration plans, including data conversion, interfaces and testing
- Establishing enterprise-level governance policies, plans, and structures to manage architecture, data and system changes related to the Integration Platform Solution
- Creating a CMS certification plan and supporting certification activities with various modules
- Establish Organizational Change Management (OCM) activities in coordination with CO
- Enterprise PMO (EPMO)
- Implementing a help desk for Operations & Maintenance (O&M) phase

Impact

- EY collaborated with Colorado to develop and maintain their MES roadmap, which involves establishing a Target Operating Model for up to 15+ modules that are being procured.
- We have also been setting up data, architectural and change control governance that will guide not only our project but all the modules as they are onboarded.
- The new solution architecture includes an enterprise service bus (ESB) to orchestrate data interfaces between systems in real time and batch, a graph database platform using blockchain technology and consent management, as well as master data management (MDM) and operational data store (ODS) capabilities for analytics and reporting.

State of Oklahoma

Oklahoma Medicaid Agency (OHCA)

Challenge

The Oklahoma Health Care Authority (OHCA) administers Oklahoma's Medicaid program, commonly known as SoonerCare. In defining its long-term Medicaid Enterprise Systems (MES) strategic roadmap, OHCA sought expand beyond the focus on technology systems to "Reimagine the Future of Medicaid" and transform its entire organization, including future implications for OHCA's products and services, processes, technology and data, and organizational capabilities and governance.

Our Solution

- Aligned cross-agency leadership on a common purpose statement and 10-year strategic roadmap of initiatives to transform the organization, including process redesign, technology and data, organization capabilities, and collaboration opportunities with related state agencies and organizations.
- Mapped end-to-end OHCA value chain and cross-constituent experience for ecosystem personas (e.g., members, providers, MCOs, CBOs, other state agencies and employees) to identify unmet needs, market, technology, and regulatory trends, and leading practices of comparable states.
- Conducted collaborative workshops and focus groups to engage OHCA stakeholders across agencies and functional processes to assess technology and data use case requirements and prioritize initiatives to build organizational capabilities over time.
- Developed OCHA's MES procurement strategy and module sequencing roadmap, defined OHCA's transformation management office (TMO), and developed an 18-month plan for implementation.

Impact

- Defined OCHA 10-year Strategic Transformation Roadmap, including prioritized initiatives across future products and services, process redesign, technology and data requirements, organizational capabilities and opportunities for cross-agency collaboration to improve enrollee outcomes.
- Defined MES technology procurement strategy for technology modularization and CMS certification.
- Engaged OHCA stakeholders and initiated culture change by developing an end-to-end constituent experience journey map and identifying cross-enterprise awareness of the OHCA operations.

Large Eastern US State Agency

Enterprise-wide Data Governance Program Creation

Challenge

The client sought EY's help in assessing and implementing a data governance program within their agency. The client attempted to implement a data governance program in the past, but none of the attempts proved to last in the long term. Through this engagement, EY focused on establishing a program with all the critical components required for sustainment including a formal governance structure, proper executive sponsorship, defined roles and responsibilities, and the appropriate change management / communication efforts to drive cultural change.

Our solution

- In Phase 1, EY assessed the client's data maturity model, analyzed their data elements, and created a program charter and roadmap initiatives.
- EY collaborated with client personnel to identify the right individuals and roles to serve as part of the data governance program.
- EY facilitated collaborative sessions with the full data governance team to establish buy-in and momentum.
- In Phase 2, EY helped to implement a data governance program that will last with the help of client leadership.
- While the program was stood up, the EY team also tackled some of the initial roadmap initiatives such as identifying priority Data Governance Policies and drafting the first set according to leading practices and organizational input.
- On the technical side, EY gathered definition, naming conventions, metadata, and source table information for 17 common data elements across 8 source systems and aligned each element to enterprise standards.
- EY evaluated patient identifiers across 8 source systems and identify considerations for the development of a Master Patient Index.

Impact

- EY efforts supported the agency in creating a sustainable approach to Data Governance.
- The team created a robust roadmap with resource/transition mapping that will allow the client to carry the momentum forward.
- The work EY delivered ultimately helped the client sustain leading data management practices, maintain higher quality and more complete data and deliver higher quality care to patients.

Government Health System

Data Governance and Operating Model Design

Challenge

A federal health agency was seeking a transformation in how data is governed and used across the enterprise. The organization was experiencing challenges in data quality, redundancies in data products and lacked centralized oversight of data management practices across the lifecycle. In addition, the use of analytics was immature, and the agency wanted to future proof itself to move toward advanced analytics and AI-driven insights. EY was engaged as a strategic teammate for a 12-month effort to objectively evaluate its data management and analytics capabilities and process against leading practices, design a data governance and operating model framework and lead implementation of a use case in one data subject area to test the governance framework, drive data literacy and demonstrate value.

Our Solution

- Canvassed the enterprise, interviewing more than 70 data producers and consumers to identify data opportunities, gaps and current process to request and develop data products.
- Leveraged an EY framework to assess maturity of data and analytic capabilities and presented findings and recommendations to inform the operating model and governance structure.
- Socialized findings with leaders to engender support for the target state design recommendations and collaborated with key IT and data stakeholders to design the target state governance structure including a Data Governance Board to provide decision authority and supporting Stewardship Council to drive data policy design and standardized use.
- Designed data governance KPIs and data management policies to support control design.
- Facilitated prioritization of two data subject areas (HR, and Patient Safety) to pilot data governance; defined stewardship roles, created data product inventory, developed data lineage and documented pain points/ opportunities for the working group to focus on.
- Quantified value of data governance pilots and socialized with leadership to garner approval of the data governance charter and proposed operating model.

Impact

- Enterprise buy-in and agreement around the proposed data governance framework to support decision governance and drive data driven decisions.
- Improved accountability and access to data in the pilot data subject area via documented product portfolio and stewardship roster.
- Clear roadmap and action plan to operationalize the data governance operating model and recommendations to mature the data catalog and centralize data request intake to standardize prioritization and resource allocation.
- Backlog of data opportunities and analytic insights that will drive value to action the target state operating model.

Appendix D – Additional Alliance Details

Additional details on relevant commercial alliance partners.

Snowflake

The following letter asserts EY's position in the Elite tier of the Snowflake Partner Network:

Snowflake, Inc.
106 East Babcock East, Suite 3A
Bozeman, Montana 59715
snowflake.com

Date: Aug 26 2025

Customer: Massachusetts Technology Collaborative on behalf of the Massachusetts AI Hub ("Customer")

Regarding: EY ("Partner")

To Whom It May Concern:

Thank you for considering Snowflake as your enterprise data cloud solution. We are happy to provide you with this letter in connection with EY's proposal and efforts in the planning phase for your Snowflake deployment.

By this letter, we confirm that as of the date above, EY is in Elite tier of the Snowflake Partner Network, the highest level of partnership for global partners that invest in their Snowflake practice area. The alliance provides EY with industry vertical resources, access to training, and guidance on Snowflake best practices. To help support the success of time critical engagements, EY can take advantage of Snowflake resources throughout the course of the engagement to provide assistance on many topics, from architecture review to ongoing support. We hope this letter provides answers to any questions you might have on our partnership.

Sincerely,

The Snowflake Alliances and Channel Team

EY | Building a better working world

EY exists to build a better working world, helping to create long-term value for clients, people and society and build trust in the capital markets.

Enabled by data and technology, diverse EY teams in over 150 countries provide trust through assurance and help clients grow, transform and operate.

Working across assurance, consulting, law, strategy, tax and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. For more information about our organization, please visit ey.com.

Ernst & Young LLP is a client-serving member firm of Ernst & Young Global Limited operating in the US.

© 2025 Ernst & Young LLP.
All Rights Reserved.

2509-11221-CS
ED None

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax or other professional advice. Please refer to your advisors for specific advice.

ey.com

Massachusetts Technology Collaborative

Grant Agreement
Standard Budget and Invoice Template
Invoiced Costs

Line Number	Budgeting																	
1	Applicant Information																	
2																		
3	Applicant: Ernst & Young U.S. LLP			Procurement No.:2026-JAII-01														
4	Budget Period:			Title of Proposed Project:														
5	Address:			Data Commons Collaborative - Discovery & Implementation Roadmap Phase 0														
6	200 Clarendon Street																	
7	Boston, MA 02116																	
8				MTC Funding Requested:														
9				Total Project Cost:														
10	Cost Elements			Total Project	Invoice 1		Invoice 2		Invoice 3		Invoice 4		Cumulative					
11					Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY					
12						Amount		Amount		Amount		Amount	Expended		Remaining			
13	I. Direct Labor																	
14	*name/title*	*HRS*	*Rate*		*HRS*		*HRS*		*HRS*		*HRS*							
15	Fixed Cost Labor			$1,673,970.00		$ -		$ -		$ -		$ -	$ -		$ 1,673,970.00			
16				$0		$ -		$ -		$ -		$ -	$ -		$ -			
17				$0		$ -		$ -		$ -		$ -	$ -		$ -			
18				$0		$ -		$ -		$ -		$ -	$ -		$ -			
19				$0		$ -		$ -		$ -		$ -	$ -		$ -			
20				$0		$ -		$ -		$ -		$ -	$ -		$ -			
21	Total Direct Labor			$1,673,970.00	0	$ -	0	$ -	0	$ -	0	$ -	$ -		$ 1,673,970.00			
22																		
23	II. Direct Labor Fringe Cost	Rate (%):		$0		$ -		$ -		$ -		$ -	$ -		$ -			
24																		
25	II. Subcontractors/Consultants/Contracted Staffing																	
26	*Firm name/title or role/type of services to be provided*	*HRS*	*Rate*		*HRS*		*HRS*		*HRS*		*HRS*							
27				$0		$ -		$ -		$ -		$ -	$ -		$ -			
28				$0		$ -		$ -		$ -		$ -	$ -		$ -			
29				$0		$ -		$ -		$ -		$ -	$ -		$ -			
30				$0		$ -		$ -		$ -		$ -	$ -		$ -			
31				$0	5	$ -		$ -		$ -		$ -	$ -		$ -			
32				$0		$ -		$ -		$ -		$ -	$ -		$ -			
33	Total Subcontractors/Consultants			$0	5	$ -	0	$ -	0	$ -	0	$ -	$ -		$ -			
34																		
35	III. Direct Materials (provide description and costs basis)																	
36	Curriculum Development & Training			$0									$0		$0			
37				$0									$0		$0			
38				$0									$0		$0			
39				$0									$0		$0			
40				$0									$0		$0			
41				$0									$0		$0			
42				$0									$0		$0			
43	Total Direct Materials			$0		$0		$0		$0		$0	$0		$0			
44																		
45	IV. Travel			$125,440									$0		$125,440			
46																		
47	V. Other Direct Costs (*list type and description such as venue rental costs and cost basis*)																	
48				$0									$ -		$ -			
49				$0									$ -		$ -			
50				$0									$ -		$ -			
51				$0									$ -		$ -			
52				$0									$ -		$ -			
53	Total Other Direct Costs			$0		$ -		$ -		$ -		$ -	$ -		$ -			
54																		
55	VI. General & Administrative Expense/Overhead	Rate (%):		$0		$ -		$0		$0		$0	$ -		$0			
56																		
57	**Total PROJECT COSTS**			**$1,799,410.00**		**$0**		**$0**		**$0**		**$0**	**$0**		**$1,799,410**			

LEGEND:

Formula cell
Data entry cell

This section is for post-contract invoicing only. Do not complete with proposal.